



CHAIRE DÉCISIONNEL CONNAISSANCE CLIENT

Introduction au data-mining

M. Gousseff-Fondation UBS

19 janvier 2017



Tentative de définition (1)

Tentative de définition (1)

Le data-mining :
prémice de la
data-science

Évolution de
l'environnement
technologique

Data mining et
statistique
décisionnelle

Décrire et prédire

Déroulement
d'une étude de
Data-Mining

Tentative de
définition (2)

Merci de votre
attention

" Le data mining est le processus général qui permet d'analyser et prévoir les comportements individuels des consommateurs à partir des immenses bases de données des entreprises."

Inspiré de G. Saporta



Tentative de définition (1)

Tentative de définition (1)

Le data-mining :
prémice de la
data-science

Évolution de
l'environnement
technologique

Data mining et
statistique
décisionnelle

Décrire et prédire

Déroulement
d'une étude de
Data-Mining

Tentative de
définition (2)

Merci de votre
attention

" Le data mining est le processus général qui permet d'analyser et prévoir les comportements individuels des consommateurs à partir des immenses bases de données des entreprises."

Inspiré de G. Saporta

Définition trop restrictive :

- ✓ Les données sont nombreuses même en dehors des entreprises
- ✓ Les applications dépassent le cadre des comportements de consommateurs

Comment peut-on alors définir le data-mining ?

Le data-mining : prémice de la data-science

Tentative de
définition (1)

Le data-mining :
prémice de la
data-science

Évolution de
l'environnement
technologique

Data mining et
statistique
décisionnelle

Décrire et prédire

Déroulement
d'une étude de
Data-Mining

Tentative de
définition (2)

Merci de votre
attention

- ✓ Données volumineuses
- ✓ Données rarement "propres"

Le data-mining : prémice de la data-science

Tentative de
définition (1)

Le data-mining :
prémice de la
data-science

Évolution de
l'environnement
technologique

Data mining et
statistique
décisionnelle

Décrire et prédire

Déroulement
d'une étude de
Data-Mining

Tentative de
définition (2)

Merci de votre
attention

- ✓ Données volumineuses
 - ✓ Données rarement "propres"
- ⇒ Bousculent les habitudes des statisticiens.

Le data-mining : prémice de la data-science



Tentative de
définition (1)

Le data-mining :
prémice de la
data-science

Évolution de
l'environnement
technologique
Data mining et
statistique
décisionnelle

Décrire et prédire

Déroulement
d'une étude de
Data-Mining

Tentative de
définition (2)

Merci de votre
attention

- ✓ Données volumineuses
 - ✓ Données rarement "propres"
- ⇒ Bousculent les habitudes des statisticiens.
- ✓ Nécessité de modélisation
 - ✓ Techniques de traitement et d'analyse complexes

Le data-mining : prémice de la data-science



Tentative de
définition (1)

Le data-mining :
prémice de la
data-science

Évolution de
l'environnement
technologique

Data mining et
statistique
décisionnelle

Décrire et prédire

Déroulement
d'une étude de
Data-Mining

Tentative de
définition (2)

Merci de votre
attention

- ✓ Données volumineuses
- ✓ Données rarement "propres"
- ⇒ Bousculent les habitudes des statisticiens.
- ✓ Nécessité de modélisation
- ✓ Techniques de traitement et d'analyse complexes
- ⇒ Dépassent les compétences traditionnelles des informaticiens.

Évolution de l'environnement technologique



Tentative de définition (1)

Le data-mining :
prémice de la
data-science

Évolution de
l'environnement
technologique

Data mining et
statistique
décisionnelle

Décrire et prédire


Déroulement
d'une étude de
Data-Mining

Tentative de
définition (2)

Merci de votre
attention

- ✓ Capacité de stockage des données
 - ✗ 1986 : Apple II c : pas de disque dur.
 - ✗ 2017 : clef USB 1 To.
- ✓ Capacité de calcul
 - ✗ Gosset à Fisher : *"Je vous envoie ces tables vu que vous êtes probablement le seul homme susceptible de vous en servir."*
 - ✗ Serveurs multi-processeurs, multicœurs.
- ✓ Diversité des algorithmes
 - ✗ Théorie Statistique
 - ✗ Théorie de l'apprentissage
 - ✗ Théorie de l'information

Data mining et statistique décisionnelle



Tentative de définition (1)
Le data-mining :
prémice de la
data-science
Évolution de
l'environnement
technologique

Data mining et
statistique
décisionnelle

Décrire et prédire
Déroulement
d'une étude de
Data-Mining
Tentative de
définition (2)
Merci de votre
attention

Situation de développement de la statistique (Fisher, Student,...) :
primat au modèle

- ✓ Données rares et chères
- ✓ Données récoltées à partir de plan d'expériences
- ✓ Priorité à l'explication des phénomènes observés



Data mining et statistique décisionnelle

Tentative de définition (1)
Le data-mining :
prémice de la
data-science
Évolution de
l'environnement
technologique

Data mining et
statistique
décisionnelle

Décrire et prédire
Déroulement
d'une étude de
Data-Mining
Tentative de
définition (2)
Merci de votre
attention

Situation de développement de la statistique (Fisher, Student,...) :
primat au modèle

- ✓ Données rares et chères
- ✓ Données récoltées à partir de plan d'expériences
- ✓ Priorité à l'explication des phénomènes observés

Situation d'émergence du data-mining : primat à la donnée

- ✓ Données abondantes
- ✓ Données peu structurées
- ✓ Priorité à la prise de décision dans un monde concurrentiel



Data mining et statistique décisionnelle

Tentative de définition (1)
Le data-mining :
prémice de la
data-science
Évolution de
l'environnement
technologique

Data mining et
statistique
décisionnelle

Décrire et prédire
Déroulement
d'une étude de
Data-Mining
Tentative de
définition (2)
Merci de votre
attention

Situation de développement de la statistique (Fisher, Student,...) :
primat au modèle

- ✓ Données rares et chères
- ✓ Données récoltées à partir de plan d'expériences
- ✓ Priorité à l'explication des phénomènes observés

Situation d'émergence du data-mining : primat à la donnée

- ✓ Données abondantes
- ✓ Données peu structurées
- ✓ Priorité à la prise de décision dans un monde concurrentiel

Décider :

- ✓ Analyses confirmatoires ET des analyses exploratoires
- ✓ Limiter la subjectivité humaine
- ✓ Traiter plus rapidement un plus grand nombre de questions



Décrire et prédire

Tentative de définition (1)
Le data-mining :
prémice de la
data-science
Évolution de
l'environnement
technologique
Data mining et
statistique
décisionnelle

Décrire et prédire

Déroulement
d'une étude de
Data-Mining
Tentative de
définition (2)
Merci de votre
attention

Décrire : mettre en évidence des infos contenues dans le jeu de données, mais masquées par le volume de données

- ✓ Liens entre des variables et représentation synthétique :
Analyses factorielles (Cours Eric Miqueu)
- ✓ Groupes d'individus :
Classification (Suite de notre cours)



Décrire et prédire

Tentative de définition (1)
Le data-mining :
prémice de la
data-science
Évolution de
l'environnement
technologique
Data mining et
statistique
décisionnelle

Décrire et prédire

Déroulement
d'une étude de
Data-Mining
Tentative de
définition (2)
Merci de votre
attention

Décrire : mettre en évidence des infos contenues dans le jeu de données, mais masquées par le volume de données

- ✓ Liens entre des variables et représentation synthétique :
Analyses factorielles (Cours Eric Miqueu)
- ✓ Groupes d'individus :
Classification (Suite de notre cours)

Prédire : extrapoler de nouvelles informations à partir de celles contenues dans les données disponibles

- ✓ Scoring
- ✓ Autres méthodes prédictives



Déroulement d'une étude de Data-Mining

Tentative de définition (1)

Le data-mining :
prémice de la
data-science

Évolution de
l'environnement
technologique

Data mining et
statistique
décisionnelle

Décrire et prédire

Déroulement
d'une étude de
Data-Mining

Tentative de
définition (2)

Merci de votre
attention

- ✓ Définition des objectifs :
À quelle question veut-on répondre ?
- ✓ Inventaire des données existantes
- ✓ Collecte des données supplémentaires
- ✓ Exploration, nettoyage, préparation des données (jusqu'à 80% du temps d'un data-scientist)
- ✓ Segmentation de la population
- ✓ Elaboration et validation des modèles
- ✓ Formation des utilisateurs au modèle
- ✓ Suivi et enrichissement des modèles

A-t-on répondu à la question ?

Tentative de définition (2)

Le data-mining, ou fouille de données, est l'ensemble des méthodes et techniques destinées à l'exploration et l'analyse de bases de données, souvent de grande taille, parfois hétérogènes, afin de détecter, de façon automatique ou semi-automatique des règles, des associations, des structures ou des tendances permettant de restituer l'information utile à la prise de décision de manière aussi parcimonieuse que possible.

Adapté de Stéphane Tufféry, Data Mining et statistique décisionnelle, L'intelligence des données, Technip

Tentative de
définition (1)

Le data-mining :
prémice de la
data-science

Évolution de
l'environnement
technologique

Data mining et
statistique
décisionnelle

Décrire et prédire

Déroulement
d'une étude de
Data-Mining

Tentative de
définition (2)

Merci de votre
attention

Merci de votre attention

Tentative de définition (1)
Le data-mining :
prémice de la
data-science
Évolution de
l'environnement
technologique
Data mining et
statistique
décisionnelle
Décrire et prédire
Déroulement
d'une étude de
Data-Mining
Tentative de
définition (2)

Merci de votre
attention



CHAIRE DÉCISIONNEL
CONNAISSANCE CLIENT

Fondation
ubs:
Université Bretagne Sud

CHAIRE D-CC "DÉCISIONNEL-CONNAISSANCE CLIENT" - IUT de Vannes - 8 rue Montaigne - 56017 Vannes Cx / 02 97 62 64 64 / chaire.d-cc@univ-ut

Vos questions sont les bienvenues.