

Faculté des arts et des sciences - Secteur des sciences

Département d'informatique et de recherche
opérationnelle

Application bancaire du data mining

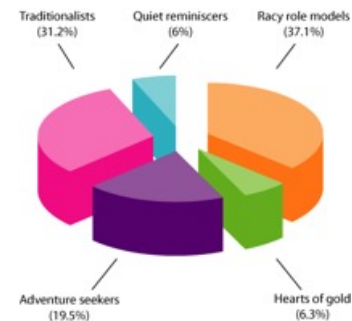
Prédiction du chiffre d'affaires

Yoshua Bengio et al., Université de Montréal

D'où vient le data mining?

Le data mining ou **fouille de données** est une des nombreuses applications commerciales de l'apprentissage statistique

Application bancaire de la fouille de données



Pourquoi est-ce intéressant?

- Parce que cela est essentiel à une mise en marché rentable et efficace tenant compte de la taille et du potentiel de croissance de l'entreprise
- Parce que cela permet d'offrir des services plus ciblés sur les besoins des clients
- Parce que le marché bancaire des entreprises est lucratif et pose des défis que le data mining sait relever

Quels sont les défis à relever?

Nos algorithmes sont spécialement conçus pour faire face aux difficultés rencontrées avec les données bancaires:

- ▣ très grand nombre de données
- ▣ valeurs manquantes
- ▣ grand nombre d'intrants potentiellement pertinents à la prédiction
- ▣ dépendances non linéaires dans les données
- ▣ intrants à grande variation

Comment s'y prendre?

Étude comparative des algorithmes suivants:

- Méthodes statistiques standard
(régression linéaire et logistique)
- Méthodes connues d'apprentissage statistique
(réseaux de neurones, arbres de décision, Adaboost)
- Nos méthodes, basées sur l'apprentissage statistique mais ajustées aux défis des données bancaires

Prédiction du chiffre d'affaires

- La priorité est de classer les entreprises en
 - très petite < 1M\$,
 - petite entreprise 1 - 10M\$
 - moyenne entreprise 10 - 100M\$.
 - À partir du profil de l'entreprise ayant 669 intrants:
 - soldes des comptes,
 - statistiques des transactions,
 - transactions bancaires
 - carte de crédit, etc
-

Entraînement d'un prédicteur

- On entraîne le modèle à partir des exemples où le chiffre d'affaires est connu

Exemple 1: intrants sur entreprise 1, chiffre d'affaires 1

Exemple 2: intrants sur entreprise 2, chiffre d'affaires 2

Et ainsi de suite sur tous les exemples.

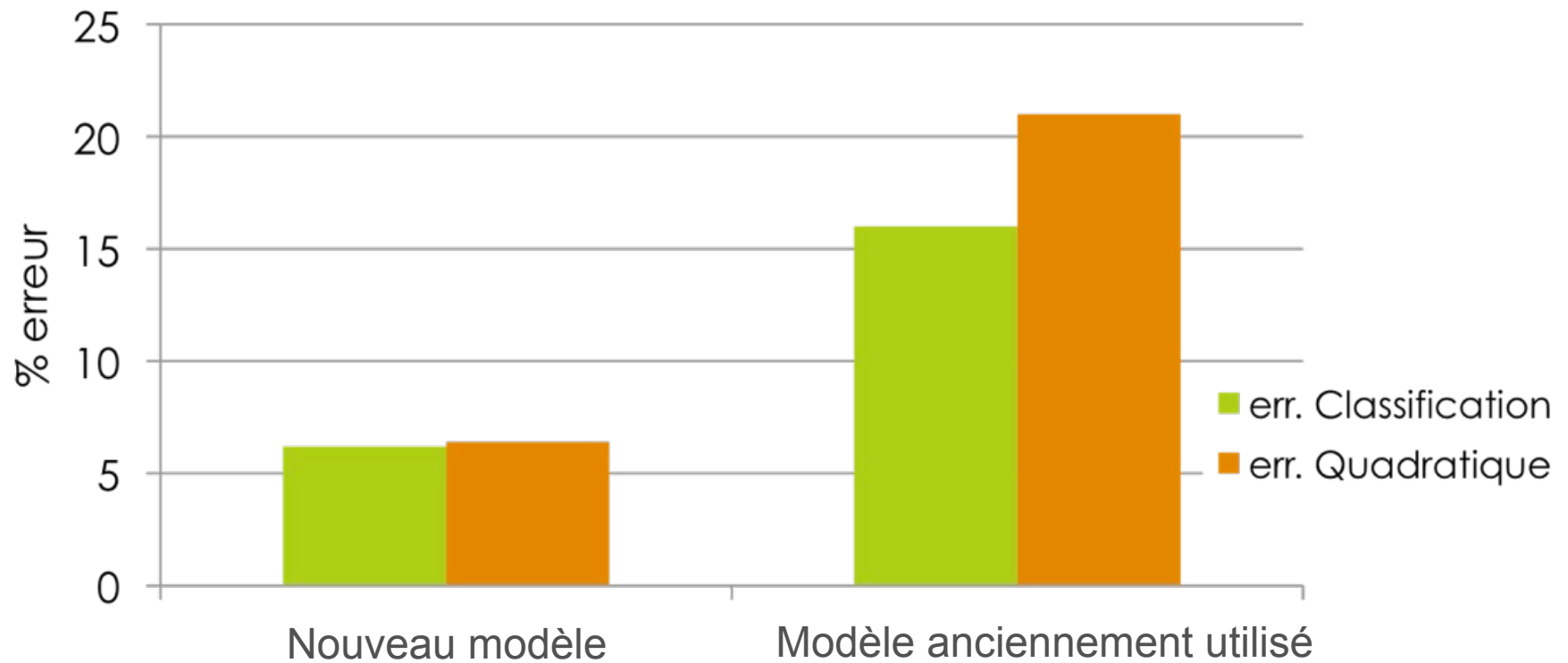


- On peut appliquer le prédicteur à des cas où le chiffre d'affaires n'est pas connu

Données bancaires

Nombre d'exemples	2008/03	2007/09	2006/09
Apprentissage	77370	70522	47057
Test	37003	36710	11765
Étiquetés	114373	107232	58822
Non étiquetés	273220	278432	328465

Performances de notre modèle



Améliorations apportées

- Entraînement + rapide (de 1 mois à 1 jour)
 - Prédiction + rapides
 - Estimation de l'importance de chaque intrant pour améliorer la confiance des utilisateurs
 - Sauvegarde automatique des traitements et du modèle
 - Stabilisation des classification d'une année à l'autre
-

Conclusion

Nous avons démontré que le modèle d'apprentissage statistique que nous proposons pour faire le data-mining est **3 fois supérieur** en performance que le modèle anciennement utilisé

Perspectives

La stratégie de développement de marché dépend de la performance des outils comme le nôtre

Notre valeur ajoutée

Nos prédictions sont des atouts pour:

- ▣ le développement des affaires
 - ▣ segmentation de la clientèle
 - ▣ marketing adapté
- ▣ la gestion de la performance
 - ▣ tableaux de bords

Yoshua Bengio, PhD



- Professeur titulaire à l'UdeM
- 17 ans d'expérience R&D industrielle (chaire CRSNG)
- Plus de 200 publications et 4300 citations à ses travaux
- Chaire de Recherche du Canada
- Récipiendaire du prix Urgel Archambault 2009
- Fellow de CIRANO et de l'ICRA
- Fondateur du laboratoire LISA depuis 1993
- <http://www.iro.umontreal.ca/~bengioy>
- VP R&D d'ApSTAT Technologies