



CETADATA

Open, Big et al Data : impact, potentiel... Et limites !

Éléments pour la conférence Asprom
3 Juillet 2013

CONFIDENTIEL

MESSAGES CLÉS

- Open, Big, géolocalisée... Vraie ébullition dans le monde des Business Data, ou tendances émergentes ?**

- Big Data : l'enjeu du prédictif – de sérieux chausse-trappes IT...
Et statistiques !**

- Open Data : opportunité ou aiguillon douloureux pour la puissance publique ?**

- la puissance de la foule : source de développement pour le big open data**



DES CONTRIBUTIONS TRÈS INÉGALES PAR MINISTÈRE, EN ABSOLU ET RELATIF

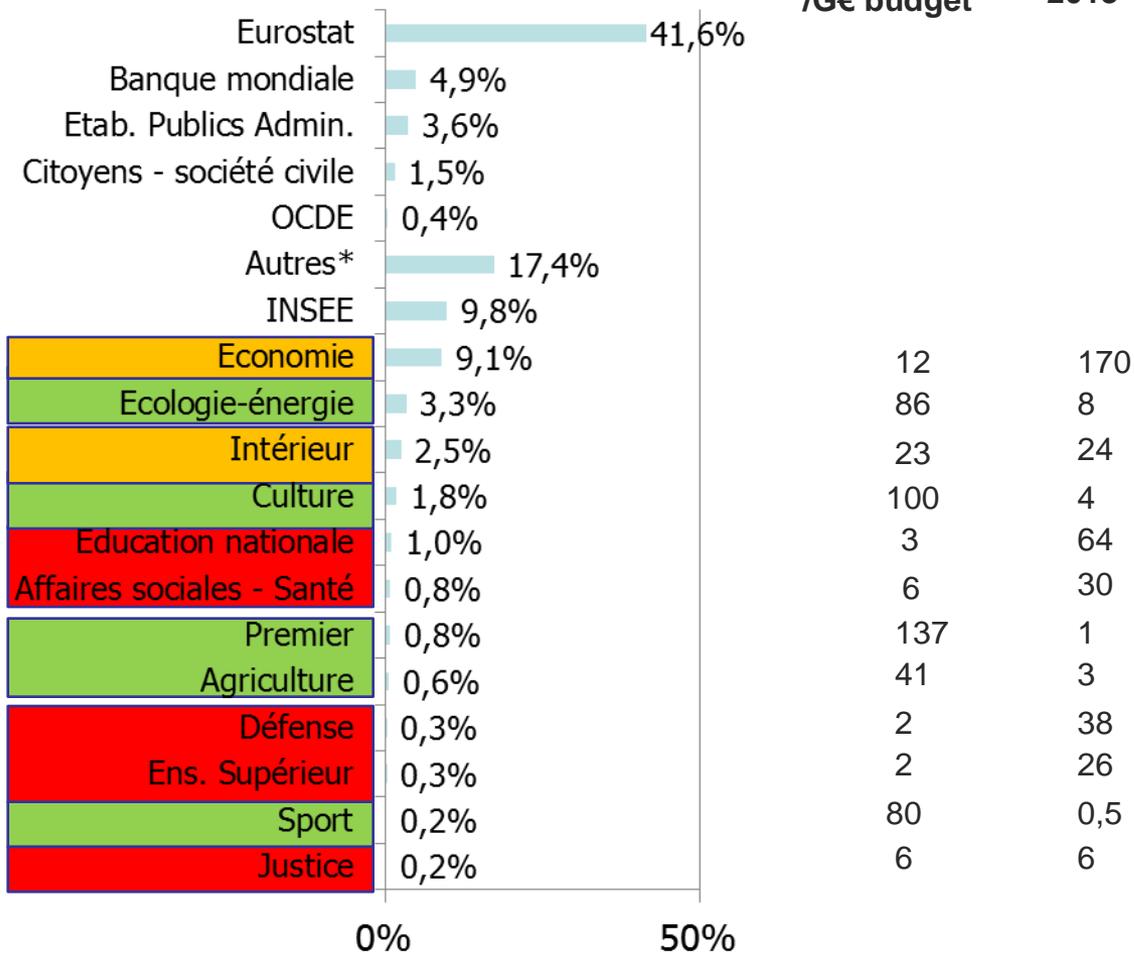
CETADATA

Organismes contributeurs

% du catalogue Data Publica

Contribution /G€ budget

Budget 2013- G€



Ministères non contributeurs

Ministère	Budget 2013- G€
Réforme Etat	?
Artisanat	?
Egalité territoires - logement	8
Affaires étrangères	3
Droit des femmes	?
Outre-Mer	2
Commerce extérieur	?
Redressement productif	3
Emploi	?

CONFIDENTIEL

* Notamment collectivités territoriales

BIG DATA : ENJEUX ET CHALLENGES DU PREDICTIF

Enjeux

- Remboursement d'un prêt
- Traitement optimal d'une maladie pour un patient donné
- Taux et chemin de transformation d'un internaute en acheteur
- Etats mésoscopiques de réseaux télécom/ énergie

Challenges

Volume des data

Nombre de tags/ identifiants

Fréquence de rafraîchissement

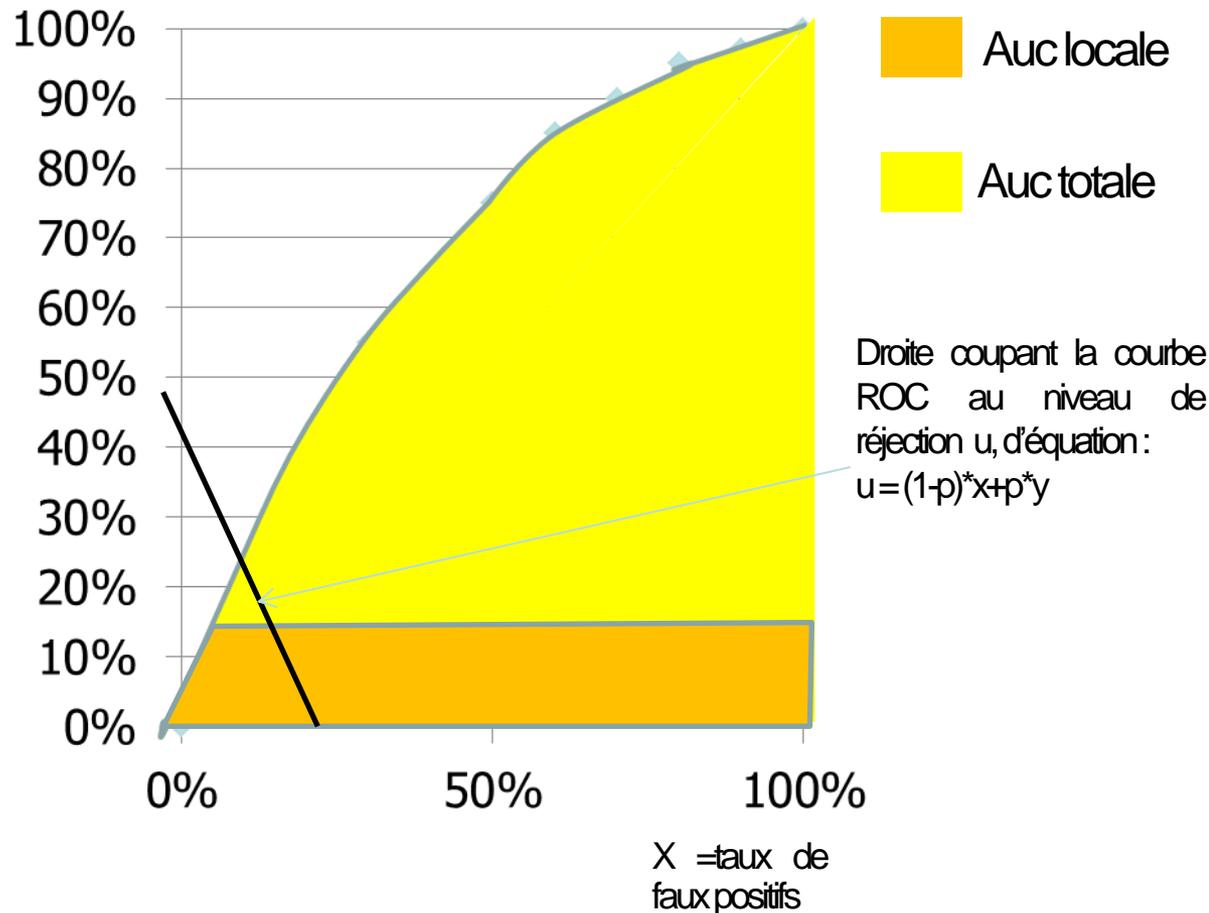
Des enjeux massifs, et des défis techniques d'origine IT et **statistique**

MESURE DE LA PERFORMANCE D'UN SCORE : AUC LOCALE...

Y= taux de vrais positifs***

Score appliqué à un phénomène présentant un taux de défaut de p ($p=10\%$ dans le cas illustré)

Recherche d'optimisation de l'auc locale correspondant aux taux de réjection** u inférieurs/proches de p



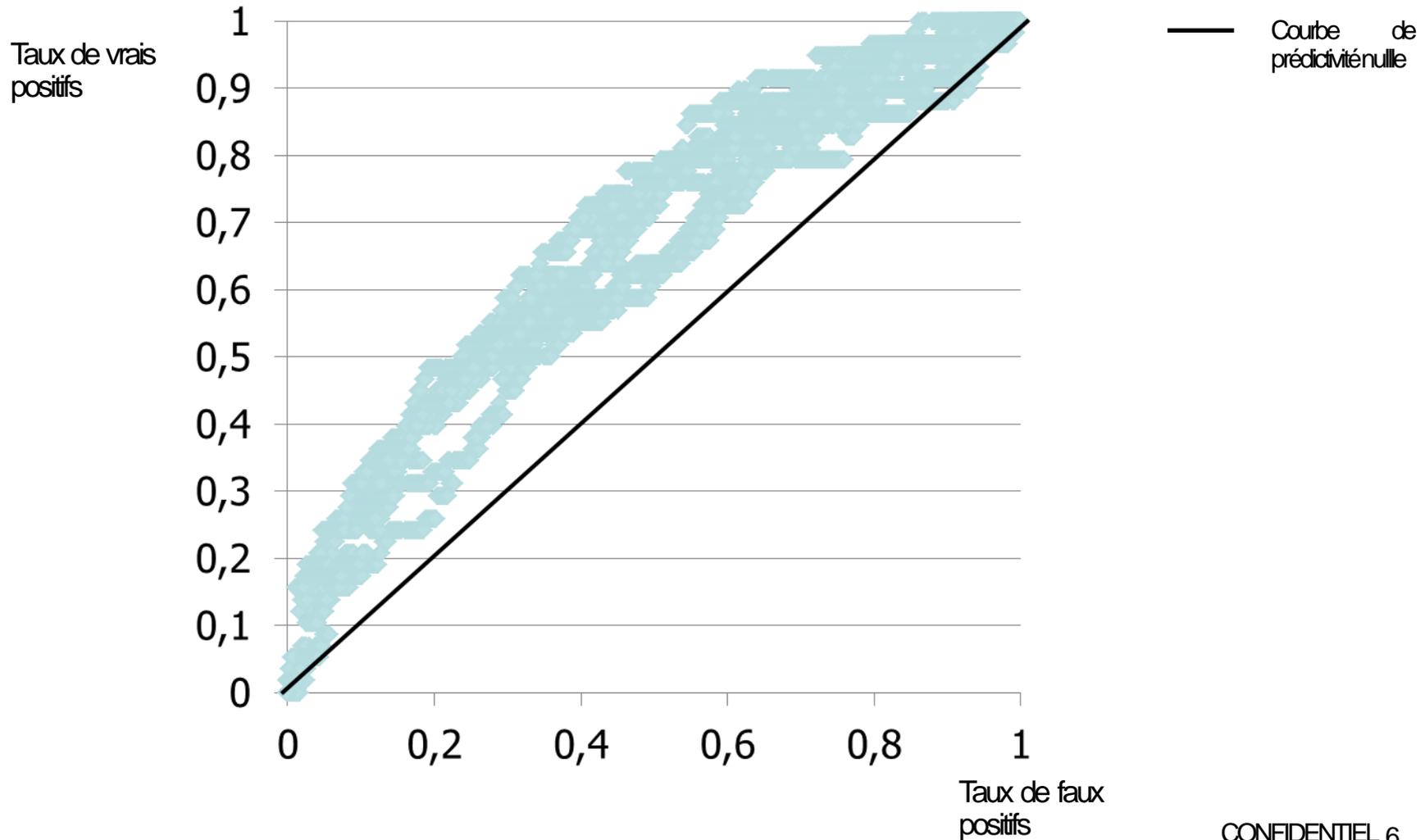
* $p= 289/2118$

** taux de réjection de $x\%$ = élimination des $x\%$ d'individus ayant le score le plus mauvais

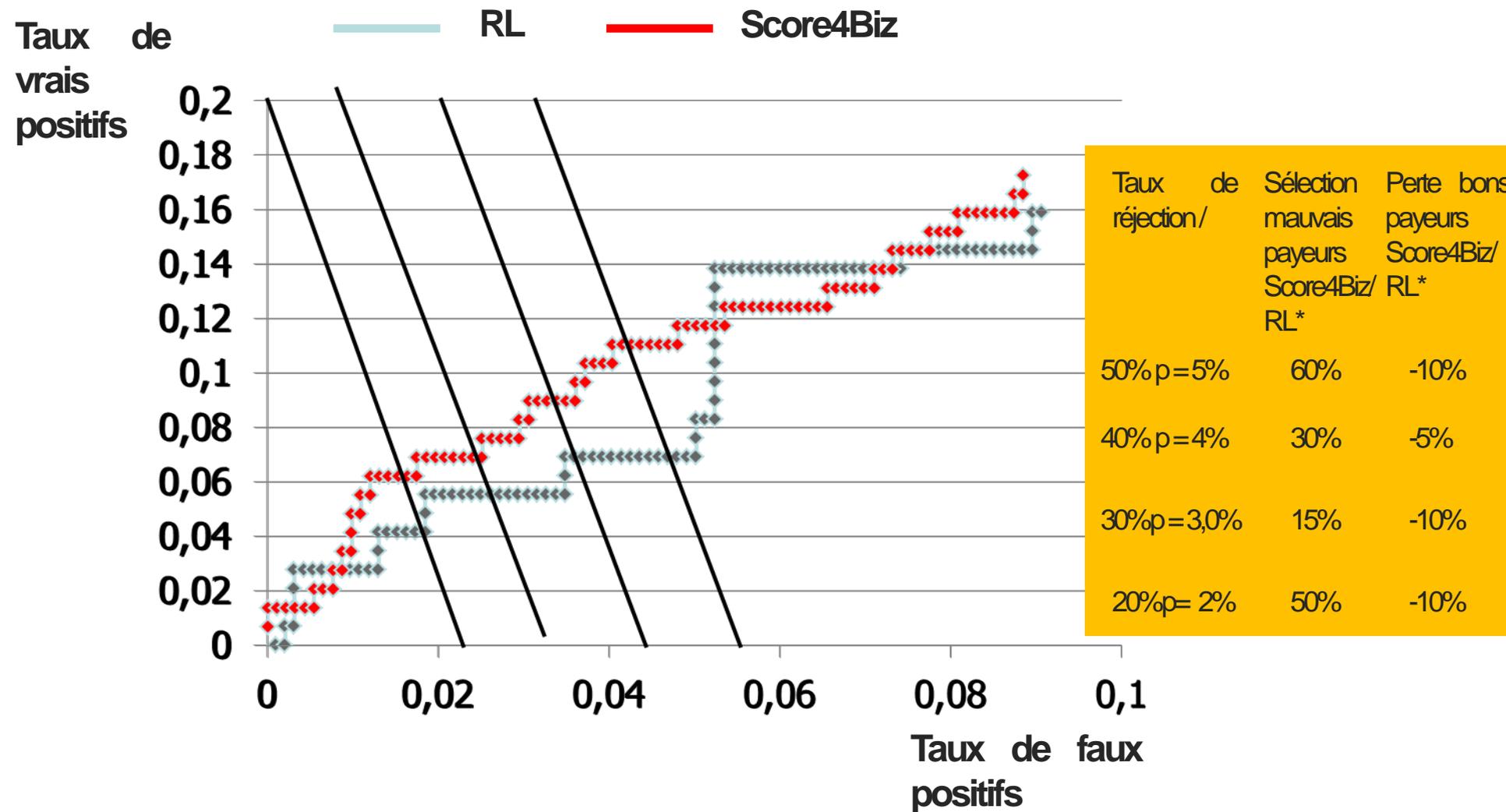
*** ici positifs = mauvais payeurs (le phénomène que l'on cherche à identifier).

Source : analyse Score4Biz

Analyse de calcul de variance d'auc pour un modèle donné de score



COMPARAISON DES COURBES ROC SCORES RL VERSUS SCORE4BIZ – LOCAL, ZONE MAUVAIS PAYEURS

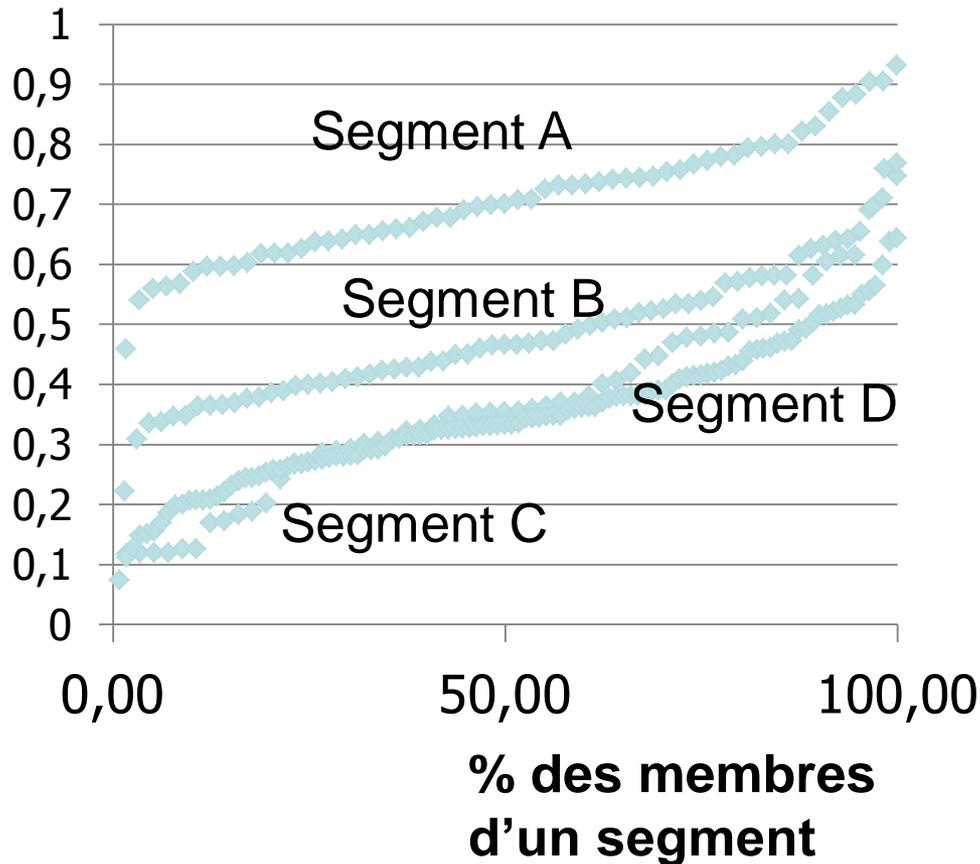


* Ratio des mauvais payeurs (resp bons payeurs) éliminés par le score Score4Biz/ bons payeurs éliminés par le score RL en éliminant les candidats ayant les scores inférieurs au score de réjection

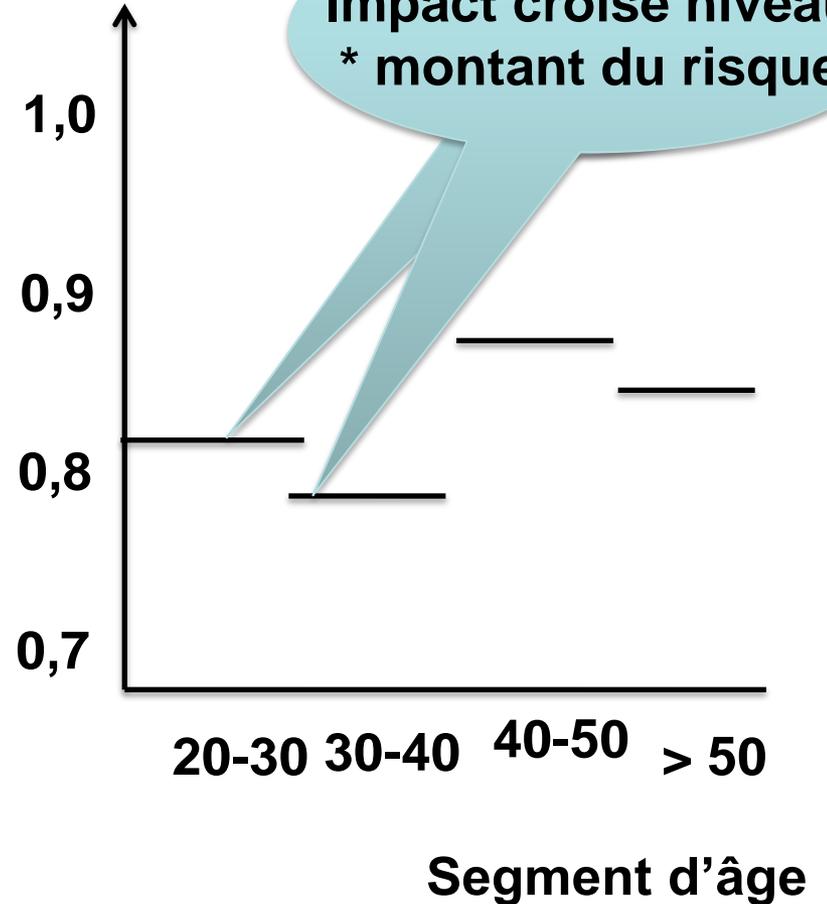
Source : analyse Score4Biz

EMPLOI DU SCORE EN LIEN AVEC LE PROFILAGE RISQUES DE SEGMENT « MARKETING »

Score de risque*



Score seuil de recrutement*



* 0 : risque très faible – 1 : risque très élevé

APPLICATION DU SCORING + SMART GOUVERNANCE : COMMENT SAUVER DES MILLIERS D'EMPLOIS ?



CETADATA

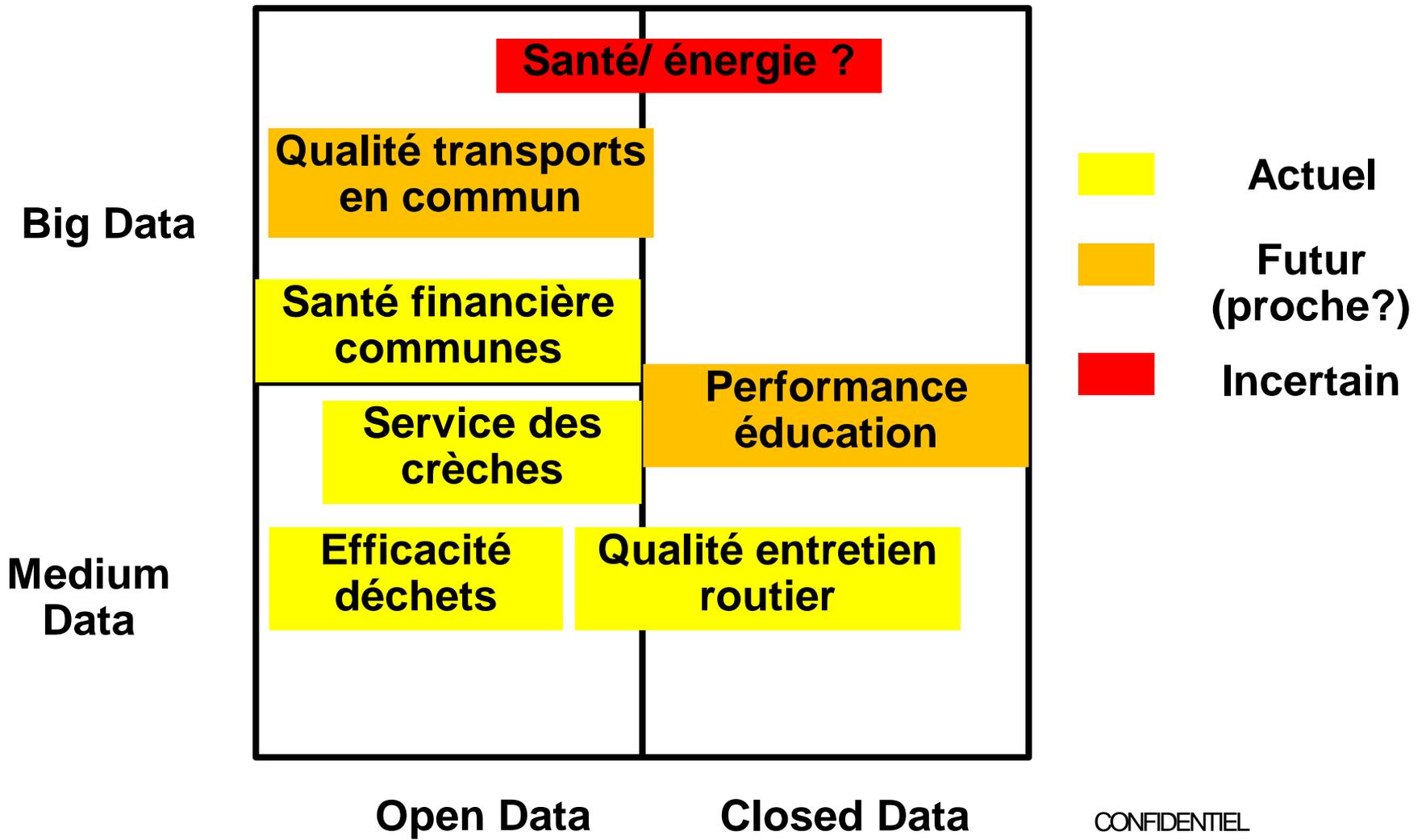
DÉTAILS LE JOUR J

CONFIDENTIEL

EXEMPLES D'APPLICATIONS À VALEUR AJOUTÉE DE L'OPEN DATA

EXEMPLES

DÉTAILS LE JOUR J



CONFIDENTIEL

SEE YOU LE JOUR J, OU À BIENTÔT AUTREMENT



F. Lainée, président

10, rue de la mairie, 95000 Boisemont

Tel : 00 33 6 14 08 26 34

Mail : f.lainee@gmail.com