

Meet up Data Business #3

4 mai 2015

Chambre de commerce des Hauts de Seine

Ordre du jour

- 18h45 – Accueil
- 19h00 – Plénière puis constitution de deux groupes
- 19h15 – Répartition en 2 groupes pour les ateliers
- 20h15 – Retour en plénière et partage sur les formulations obtenues
- 20h40 – Echange sur la méthode et les moyens de l'améliorer
- 21h00 – Pot / networking



Pourquoi sommes nous là

- La data n'est pas magique : on ne peut pas faire l'économie de réfléchir à un objectif précis dès le début du projet
- La démarche déroulée en atelier n'a rien de nouveau : vous le faites déjà au moindre ciblage.
- La méthode n'est pas finalisée : nous allons l'améliorer ensemble !

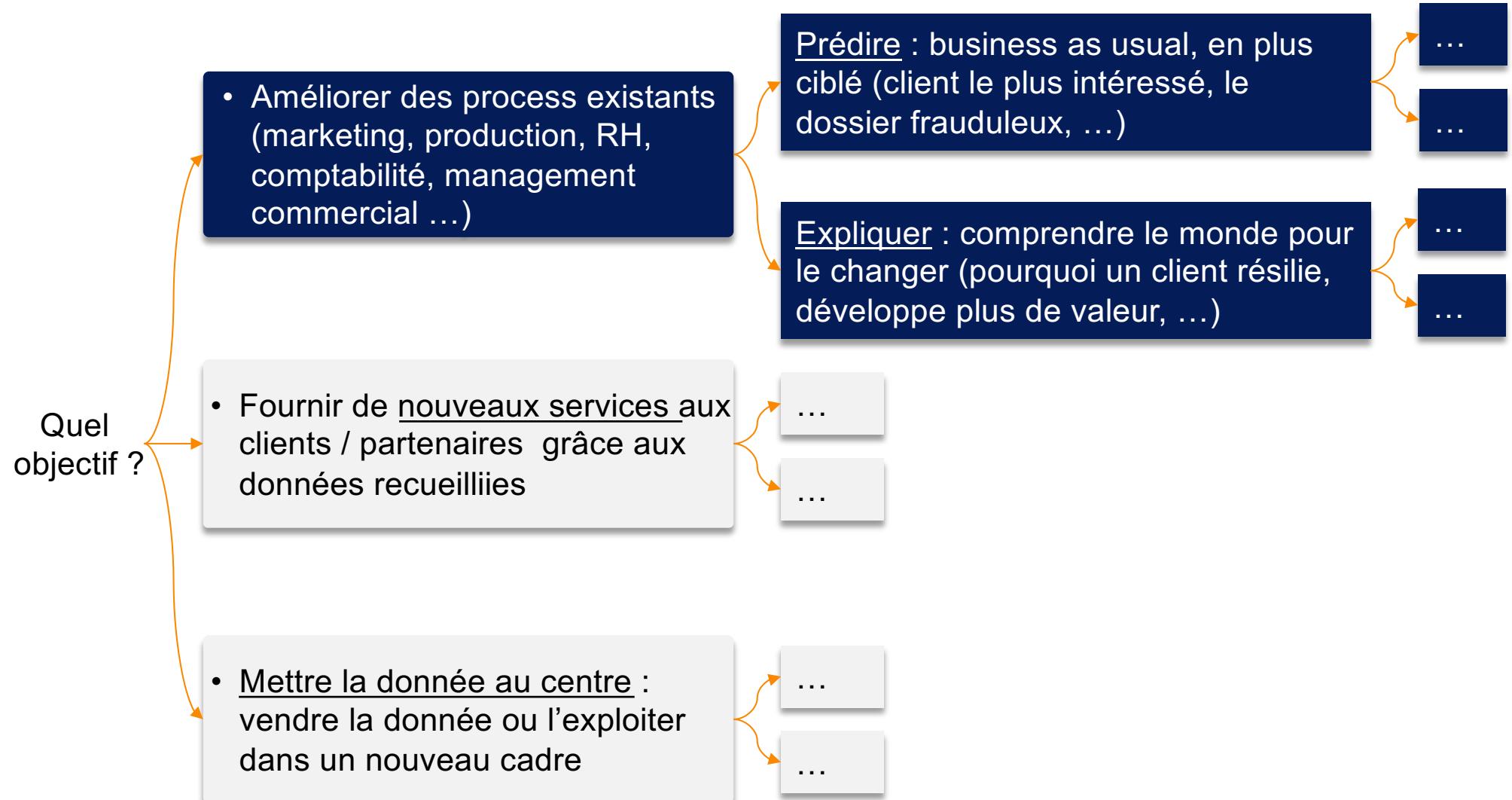
Bien formuler une question data : l'enjeu principal

Parce que c'est la phase initiale qui conditionne la suite

Parce que peu documentée et outillée, contrairement à la partie analytique elle même

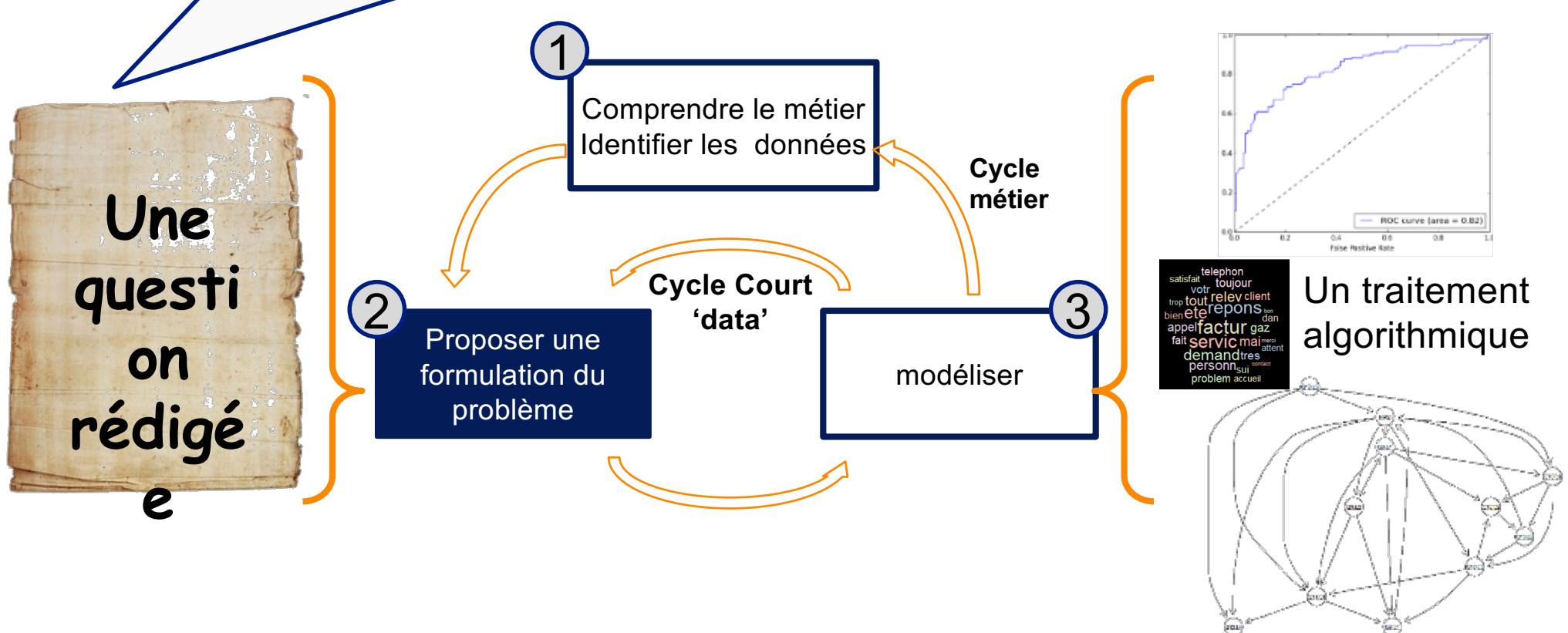
The screenshot shows the homepage of L'Usine Digitale. At the top, there's a navigation bar with links for 'TOUTE L'INFO', 'L'USINE NOUVELLE', 'INSCRIVEZ-VOUS À LA NEWSLETTER', 'DIGITAL AVENUE', and a search bar. Below the navigation, there's a main article titled 'Organiser un concours de datascientists en 5 étapes'. The article includes a sub-headline 'En 2014, la SNCF a lancé un modèle prédictif de la fréquentation de la ligne Transilien', a sidebar with 'A LIRE SUR LE M...', and a right-hand sidebar with 'Les plateformes qui proposent des challenges aux datascientists sont de plus en plus utilisées par les entreprises. Mode d'emploi'. A large callout box on the right is titled '2. BIEN FORMULER LA QUESTION' and contains text about defining a metric for a data challenge. At the bottom of the page, there's a URL: <http://www.usine-digitale.fr/editorial/organiser-un-concours-de-datascientists-en-5-etapes.N326894#xtor=EPR-4>.

A quoi sert la donnée ?

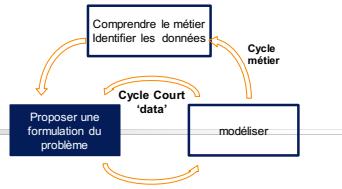


Objectif du jour : s'entraîner à formuler un problème métier pouvant être éclairé par un algorithme

Le big data n'est pas un outil de décision stratégique mais une stratégie de la micro décision



Comment reconnaît on une bonne formulation ?



3 critères pour une bonne formulation

Compréhensible pour un algorithme

vous parlez à une machine : levez toutes les ambiguïtés

Bien éclairée par vos données actuelles

Enrichissez progressivement les données utilisées : vous en découvrirez régulièrement de nouvelles

Déclenchant une action ayant un impact métier significatif

En clair : ca marche !

Niveau de maîtrise au début du projet



By design

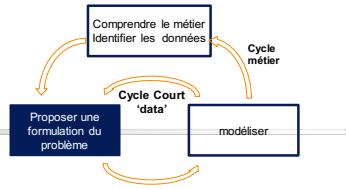
Tester à l'écran

Tester en vrai



Tous les paramètres de la prise de décision seront amenés à évoluer au fil de votre apprentissage

Formulation : un algorithme ne réponds qu'à des questions extrêmement précises



Un exemple ?

Définir un premier cadre

Levers les ambiguïtés

Fusionner en une formulation

Le comportement

- ▷ Des clients churnent chaque mois

L'action

- ▷ Envoyer une offre promotionnelle « avant qu'ils ne soient partis »
- ▷ churn ? : 3 semaine d'inactivité par exemple

- ▷ Poster un mail offre promotionnelle vers les clients qui vont devenir inactif dans les 2 mois qui viennent

Identifier parmi mes clients actuels ceux qui ont le plus fort risque de démarrer une inactivité d'au moins 3 semaines dans les 2 prochains mois

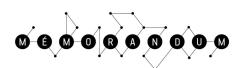
La probabilité que vous trouviez du premier coup une formulation efficace (compréhensible pour un algorithme, bien éclairé par vos données actuelles (bon score de prédiction) et induisant un bon impact business ... est faible

Et maintenant Au travail !

➤ Pour constituer 2 groupes ..

- Qui a déjà un sujet à traiter (partir d'un cas réel) ?
- Qui est professionnel de la donnée ?

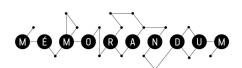
Objectif : mixer
les questions et
les réponses !



Welcome back

➤ Chaque groupe présente en 5 min

- Le sujet tel que formulé au début (améliorer mon process comptable «)
- La question formulée in fine (« détecter les factures de plus de X€ qui seront en impayé dans 2 mois »)
- Les problèmes rencontrés



Bilan

- Que pensez vous de l'exercice ?
- Qu'est ce qui permettrait d'améliorer l'approche ?
- Les possibilités d'exploitation de la donnée vous paraissent elles plus claires ?

