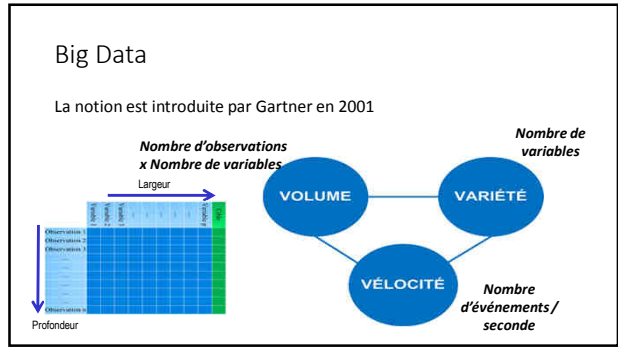
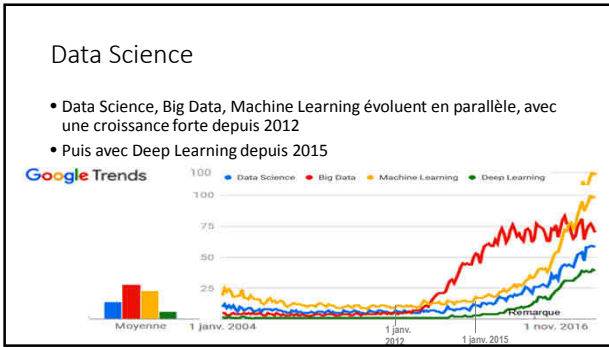


- Agenda
- Qu'est ce que la Data Science
 - Le process du projet Data Science
 - Quelques exemples de terrain
 - Devenir *data-driven*

Qu'est ce que la Data Science

Data Science

- La Data Science est un nouveau domaine, ensemble de techniques et outils, qui ont émergé pour traiter les Big Data
- La Data Science
 - Est centrée métier
 - S'appuie sur les données
 - Fait appel à de nombreuses techniques
 - Machine learning & statistiques
 - Informatique
 - Nouvelles architectures (Hadoop, Hdfs, Spark, Cloud...)
 - Nouvelles bases de données (NoSQL)
 - Nouveaux outils (open-source)



Le Big Data est possible aujourd'hui

Grâce aux données et à un marché d'outils logiciels très complet

L'Intelligence Artificielle est critique pour l'économie

AI = Big data + Machine Learning + Deep Learning = Data Science 2.0

France L'intelligence Artificielle va désormais jouer un rôle bien plus important que celui qu'elle jouait jusqu'alors. Elle n'est plus seulement un programme de recherche confiné aux laboratoires ou à une application précise. En effet, nous sommes dans un monde numérique... un monde de données... Dans ce monde là... ces technologies déterminent notre capacité à organiser les connaissances, à leur donner un sens, à augmenter nos capacités de prise de décision et de contrôle des systèmes. Et notamment à tirer de la valeur des données. L'intelligence artificielle est donc une des clés du pouvoir de demain dans un monde numérique.

US Artificial intelligence (AI) is a transformative technology that holds promise for tremendous societal and economic benefit. AI has the potential to revolutionize how we live, work, learn, discover, and communicate. AI research can further our national priorities, including increased economic prosperity, improved educational opportunities and quality of life, and enhanced national defense and security. Because of these potential benefits, the U.S. government has invested in AI research for many years.

China The rapid development of artificial intelligence (AI) will profoundly change human society and life and change the world. To seize the major strategic opportunity for the development of AI, to build China's first-mover advantage in the development of AI, to accelerate the construction of an innovative nation and global power in science and technology in accordance with the requirements of the CCP Central Committee and the State Council, this plan has been formulated.

- ### Exemples d'applications à la connaissance client
- Améliorer le ciblage des campagnes marketing
 - Développer les marchés de niche
 - Produire de meilleurs modèles de prévision
 - Pour la résiliation, la détection de la fraude ...
 - Développer les nouvelles plateformes de distribution
 - Améliorer l'interaction client (NBA, temps réel)
 - Proposer des recommandations
 - Concevoir de nouveaux produits
 - Ex: Usage-based-insurance
 - Connected xx (car, health, home)
- Les données peuvent aider les compagnies d'assurance à mieux cibler leurs campagnes publicitaires et à mieux comprendre leurs clients.*
- Pourquoi les objets connectés font rêver les compagnies d'assurances**
-

Le process du projet Data Science

Le process du projet Data Science

- Les étapes sont classiques
- Mais avec quelques grosses différences avec des big data
 - La génération de features (variables)
 - L'approche *Machine Learning* ≠ Statistiques
 - La construction du modèle
 - Le plus important: les données
 - L'utilisation d'infrastructures IT différents (Hadoop, Spark)
 - La *programmation* en plus des statistiques
 - La programmation est très massivement nécessaire pour les étapes de préparation des données (80% du temps passé dans un projet)
 - Nouveaux langages: Python, ou R

L'impact des Big Data

Modèles prédictifs: Plus les données sont "Big" mieux c'est:

- Augmenter le volume
- Augmenter la variété – feature engineering

Data Science

Règles absolues

- Se focaliser sur les données
 - Puis, si on a le temps, sur l'algorithme
- Toujours augmenter le volume des données
 - Ne jamais éliminer des données a priori, avant d'avoir construit un modèle
- Toujours augmenter la variété
 - Ne jamais éliminer des variables a priori, avant d'avoir construit un modèle
 - Ajouter de nouvelles sources de données
 - Image, parole, texte, données sociales
 - Ajouter de nouvelles variables « dérivées »
 - Feature Engineering

Quelques exemples de terrain

La fraude à la carte bancaire sur Internet

- La vente en ligne augmente très rapidement
- La fraude aussi
- Difficultés
 - Taux de fraude très faible
 - <<1%
 - Taux d'alerte très faible
 - <<1%
 - Volumes très grands
 - La fraude change très vite

Banque de France, Rapport annuel OSCP, 2012

La fraude à la carte bancaire sur Internet

- Méthode : créer des variables additionnelles (Variété ↑)
 - 37 → 997 (~1500 avec scores cartes & marchands + Agrégats)

- Variables Sociales
- Et segmenter

La fraude à la carte bancaire sur Internet

- Résultats

Modèle	Précision	Rappel
Baseline	8,18%	1,40%
Baseline + Agrégats	19,00%	9,13%
Baseline + Agr. + Social	40,58%	9,09%
Seg 19	28,21%	5,09%
Seg 19 + Agr.	28,82%	7,38%
Seg 19 + Agr. + Social	60,89%	16,46%
- Importance des variables
 - Variables Initiales
 - Agrégats Marchand / Agrégats Carte
 - Variables Sociales: Carte / Marchand


Recommandation de contenu – Mondomix

- Lors de la consultation d'un article, proposer d'autres contenus
 - Valorisation du catalogue d'articles
 - Augmentation du temps passé
- Méthodes
 - Utilisation des contenus : Réseau Contenus x Tags
 - Utilisation des logs : Réseau Contenus x Cliquezurs



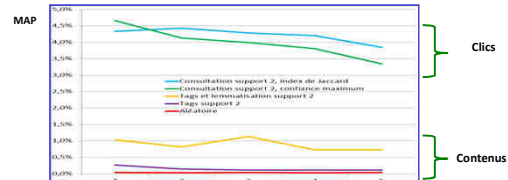
Recommandation de contenu – Mondomix

- Utilisation du contenu
 - Extraction de mots permettant de caractériser un contenu



Recommandation de contenu – Mondomix

- Comparaison des méthodes
 - Recommandations basées sur les contenus
 - Recommandations basées sur les consultations (clics)




Le quart d'heure chinois


- La Chine découvre la grande consommation
 - La Relation Client bénéficie de l'effet « dernier arrivant »
 - Tout se fait par téléphone (« relevé bancaire »)
- Exemples
 - WeChat offre « tout » au client
 - JD.com arrive aux pieds du client



WeChat

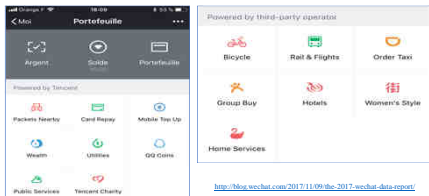
- Tencent launched WeChat in 2011 as a messaging app
- It is the largest social network, platform for payments, games...
 - 5th largest world internet company
- 980 M monthly active users in Q3 2017.
- Soon will be an electronic ID system



Tencent 腾讯 

WeChat

- A portfolio of multiple apps on WeChat
 - Over 20+ industry categories and 200+ sub-industry categories covered



<http://www.wechat.com/2017/11/09/06-2017-wechat-data-report>

JD.COM 京东 多快好省

JD.com

- Créé en 2014, retail B2C
- 3^{ème} société internet mondiale (devant Facebook et Alibaba)
- 266,3 million d'utilisateurs actifs en septembre 2017
- JD.com vise l'expansion en France
 - Exemple: 2 accords signés en janvier au Forum IA à Pékin
 - Business France and JD.com for implantation in France of JD.com and increasing French products presence on JD.com platform (2 B € sales planned)
 - Agreement framework between Fives and JD.com for logistics (100 M€).



JD.COM 京东 多快好省

JD.com

- « door to door » « last mile delivery »
 - Dans les villes

Les campagnes



<https://www.cnn.com/video/2017/06/18/how-e-commerce-uses-drones-to-deliver-to-the-outskirts-of-china.html>
<http://ummanedatag.org/chinese-delivery-drones>

Devenir *data-driven*

Une démarche agile

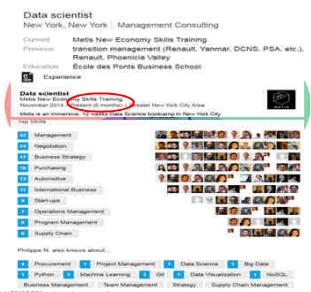
- S'appuyer sur une équipe de direction soudée autour des objectifs des actions Data Science
 - Certaines entreprises créent une équipe Data Science dédiée
 - Centre d'excellence ou Lab
- Choisir les cas d'usage
 - Pour démontrer rapidement la méthode & l'apport de valeur pour l'entreprise ET le client
 - Passer rapidement à la mise en production
- Prévoir les phases de sensibilisation / formation dans toute l'entreprise
 - La donnée doit devenir un actif sur lequel chacun a une part de responsabilité
 - La question du Change management est cruciale pour le succès

Un processus avec des étapes clés

- Comprendre les enjeux et la valeur
- Réunir une équipe
- Mettre en œuvre l'infrastructure nécessaire
- Réaliser les premiers cas d'usage & démontrer la valeur
- Passer des phases pilotes à la mise en production
- Diffuser dans les entités
- Former / re-skiller dans toute l'entreprise

Réunir une équipe

- Les compétences requises
- Les points critiques
 - Recrutement
 - Ressource rare
 - Attention aux CV !
 - Formation
 - Re-skilling



Data Scientist: The Sexiest Job of the 21st Century
Harvard Business Review

<http://www.hbr.org/2012/01/the-data-science-ven-digman>

Questions ?