Dessine moi le Big Data

#LOIRetCHERTECH

millef'oeil

Plan

1- Introduction
2 - Le Big Data c'est quoi ? Mythe ou réalité ?
3 - L'histoire et la genèse du Big Data
4 – Les applications majeures du Big data
5 - Les uses cases dans le marketing (Ciblage et calcul du Churn)
6 – L'éthique dans tout cela ? La France protectionniste ?

Plan

1- Introduction
2 - Le Big Data c'est quoi ? Mythe ou réalité ?
3 - L'histoire et la genèse du Big Data
4 – Les applications majeures du Big data
5 - Les uses cases dans le marketing (Ciblage et calcul du Churn)
6 – L'éthique dans tout cela ? La France protectionniste ?

Christophe Cerqueira

- *Directeur Pôle Big Data Groupe Cyres*
- *Maitre Conférence Université de Tours*
- *Membre conseil de surveillance pour l'obtention du titre d'ingénieur Polytech Tours*
- *Rédacteur au journal du Net*

Mob: 06.14.51.62.48

ccerqueira@cyres.fr

CYRÈS Groupe

Database & Big Data

INGENSI

Cloud & Datacenter

HOSTEAM

Collaboratif & Digitale

INTERACTIVE

Plan

1- Introduction
2 - Le Big Data c’est quoi ? Mythe ou réalité ?
3 - L’histoire et la genèse du Big Data
4 – Les applications majeures du Big data
5 - Les uses cases dans le marketing (Ciblage et calcul du Churn)
6 – L’éthique dans tout cela ? La France protectionniste ?

INGENSI

Rappels sur les mesures des données

Préfixe			Peu usitée
Yotta	X 1 000 000 000 000 000 000	10^{24}	Quadrillion
Zetta	X 1 000 000 000 000 000 000	10^{21}	Trilliard
Exa	X 1 000 000 000 000 000 000	10^{18}	Trillion
Péta	X 1 000 000 000 000 000	10^{15}	Billiard
Téra	X 1 000 000 000 000	10^{12}	Billion
Giga	X 1 000 000 000	10^{9}	Milliard
Méga	X 1 000 000	10^{6}	Million
Kilo	X 1 000	10^{3}	Mille

2- Le Big Data C'est Quoi ? , solution au data déluge ?

Les systèmes actuels sont incapables de gérer de telles quantités :

80%
de l'information est « non-structurée » [4]

95%
de l'information est non-exploitée [4]

…Début d'une nouvelle ère

(1) IDC, 2011 - (2) Gartner, 2011 - (3) Radicati Group, 2009 - (4) Forrester, 2011

Quelques chiffres

- 3,8 zettaoctets en 2015

soit une pile de blu-ray qui ferait 7 fois le tour de la Terre

- 60% de croissance/an des volumes d'informations mais que 5% pour les budgets informatiques
- Un Boeing produit 20 To/heure de données
- 250 milliards d'emails envoyés par jour (80 % de spam)
- 72h de vidéos déposées par minute sur Youtube

préparer et anticiper ? Sans Doute

- Les solutions actuelles répondent mal (pas) aux problématiques liées, avec un TCO élevé (Exadata d'Oracle, Netezza d'IBM, etc.)

- Les applications doivent changer
  - Dimensionnées à l'échelle de la planète
  - Flux de données complexes, multiples et en temps réel
  - Agilité à tous les niveaux : analyse, stockage, restitution

- Pour
  - tirer profit de ses données mais également de celles qui sont à portée de main,
  - répondre à des besoins qui pour le moment n'étaient pas adressables

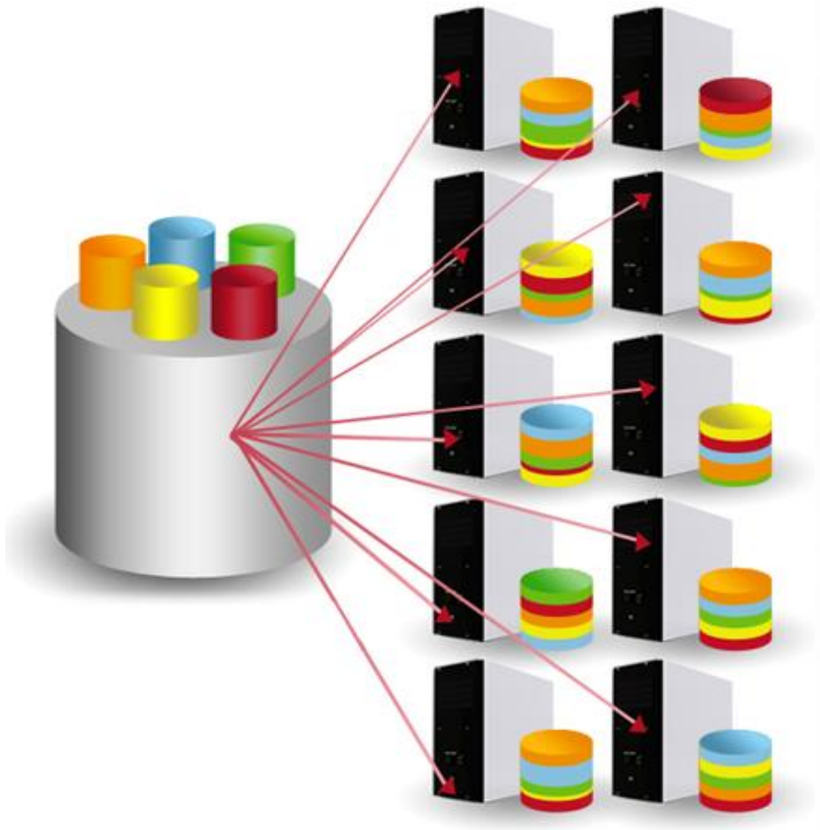
- … et tout ça en temps réel

Big Data et croissance des données

Plan

1- Introduction
2 - Le Big Data c'est quoi ? Mythe ou réalité ?
3 - L'histoire et la genèse du Big Data
4 – Les applications majeures du Big data
5 - Les uses cases dans le marketing (Ciblage et calcul du Churn)
6 – L'éthique dans tout cela ? La France protectionniste ?

3 – L'histoire : Les systèmes distribués de données



HDFS

- Objectifs et principes de GFS/HDFS

✓ Stocker de grandes quantités d'information à moindre coût » utilisation de serveurs courants (Idéalement des fichiers volumineux)

✓ Haute tolérance aux pannes » donnée répliquée 3 fois sur 3 serveurs géographiquement distants

✓ Stockage extensible à volonté » ajout à chaud de serveurs pour augmenter les capacités de stockage et de traitement de l'architecture

3 - L'histoire : Google Le système de fichier GFS

- Pour stocker son Index Grandissant quelle solution pour Google ?

Utilisation d'un SGBDR ?

➔ Problème de distribution des données

➔ problème du nombre d'utilisateurs

➔ problème de Vitesse du moteur de recherche

- Invention d'un nouveau système Propriétaire : GFS (Google File Système en 2003)

3 - L'histoire : le Big Data, Google MAP/REDUCE

INGENSI

Google

- La notion de Big Data est intimement lié à la capacité de traitements de gros volumes ➔ Un nouvel Algorithme a été mis au point….
- Le premier Article a été publié en 2004 : Jeffrey Dean and Sanjay Ghemawat

MapReduce : Simplified Data Processing on Large Clusters

- C'est un algorithme inventé par Google, Inc afin de distribuer des traitements sur un ensemble de machines avec le système GFS
- Google possède aujourd'hui plus de 2 000 000 de serveurs interconnectés dans le monde

3 - L'histoire : le Big Data, Google et les autres

- ✓ Contributeur de l'implémentation Libre du système GFS

  (Dugg Ketting)

Les pures players de l'internet ont choisi d'utiliser ces algos distribués. (HDFS et MAPREDUCE)

- Facebook
- Twitter
- LinkedIn
- ….

Le cas Facebook
(L'échelle de la planète : 1 000 000 000 d'utilisateurs)

- Plus de 500 To de données par jour , soit l'équivalent du contenu de 20 000 disques Blu-ray simple-couche transitent chaque jour par les serveurs de Facebook.
- 2,7 milliards de boutons « J'aime » sont cliqués par jour
- 300 millions de photos sont chargées par jour en moyenne (2 milliards en Pic)
- En conséquence, les serveurs du réseau social doivent analyser en moyenne toutes les demi-heures l'équivalent de 105 To de données.
- Facebook dispose d'un cluster distribué de 100 Po

Plan

1- Introduction
2 - Le Big Data c'est quoi ? Mythe ou réalité ?
3 - L'histoire et la genèse du Big Data
4 – Les applications majeures du Big data
5 - Les uses cases dans le marketing (Ciblage et calcul du Churn)
6 – L'éthique dans tout cela ? La France protectionniste ?

4 – Les applications majeures du big Data

Big Data pour moi c'est quoi ?

- Effet de mode ou révolution?
  - ~~✓ capacité à stocker des pétaoctets sur 1 an~~
  - ✓ capacité à traiter des téraoctets en 1 minute
  - ✓ **La donnée comme source de profit**

    non pas comme un coût

4- Identifier un projet Big Data : les 3V

- ✓ **Volume**
  Saturation des systèmes actuels avec toujours plus de données

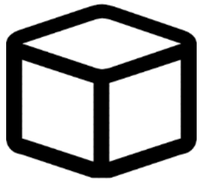
- ✓ **Vélocité**
  Quel délai pour prendre une décision à partir de l'information collectée ?

- ✓ **Variété**
  Intégrer une multitude de formats différents provenant d'une multitude de sources de données

4- Identifier un projet Big Data : le 4 ème V

✓ Valoriser les données

Fournir des outils innovants aux décideurs , aux utilisateurs , aux Partenaires



Exploiter les données

Mesurer

Indexer/Visualiser

4 – Pourquoi le Temps Réel dans le Big Data ?

« Analyser les informations en Temps réel »

Productivité
Agilité des architectures
Amélioration de la qualité

Services
Innovation
Centre de profits
Nouvelles applis

Indicateurs
Nouvelles corrélations
Connaissance de l'instant

4 – Dans quel secteur pour le Big Data ?

« Tous les secteurs sont concernés »

- Santé et imagerie médicale
- Optimisation des couvertures des services de secours
- Analyse des épidémies

- Reconnaissance faciale
- Villes connectées
- Optimisation des chaines de Productions
- Optimisation de la maintenance

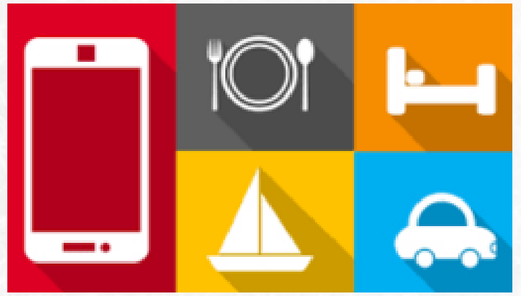
- Connaissance du client
- Connaissance de l'instant d'achat
- Conception des produits
- Adaptation de l'offre et de la demande en temps réel

Plan

1- Introduction
2 - Le Big Data c'est quoi ? Mythe ou réalité ?
3 - L'histoire et la genèse du Big Data
4 – Les applications majeures du Big data
5 - Les uses cases dans le marketing (Ciblage client et calcul du Churn)
6 – L'éthique dans tout cela ? La France protectionniste ?

5- Use case N°1: L'assureur connecté ….

Un grand assureur veut équiper ses clients qui ont une assurance Voiture d'un boitier connecté. (Pay as You Drive)



- Quel est son but premier ?
  - Faire payer les bon conducteurs moins cher ?
- Que va-t-on faire des données ?
  - Faire des statistiques pour afficher les horaires d'utilisations ?

5- Un use case pour bien comprendre !! L'assureur

Un grand assureur veut équiper ses clients qui ont une assurance Voiture d'un boitier connecté. (Pay as You Drive)

- Quel est son but premier ?

  - Diminuer les sommes à rembourser en faisant du profiling et de l'analyse comportementale pointue
  - Augmenter ses marges….

- Que va-t-on faire des données ?

  - Faire du prédictif pour analyser les mauvais conducteurs
  - Calculer une surtaxe sur les clients à risques
  - Vendre vos habitudes de « roulage » à des enseignes

5- Un use case pour bien comprendre !! L'assureur

Calculons le volume !!!!

3 000 000 Objets connectés

10 captages / mn

Historisation sur 4 ans

20 Go / heure

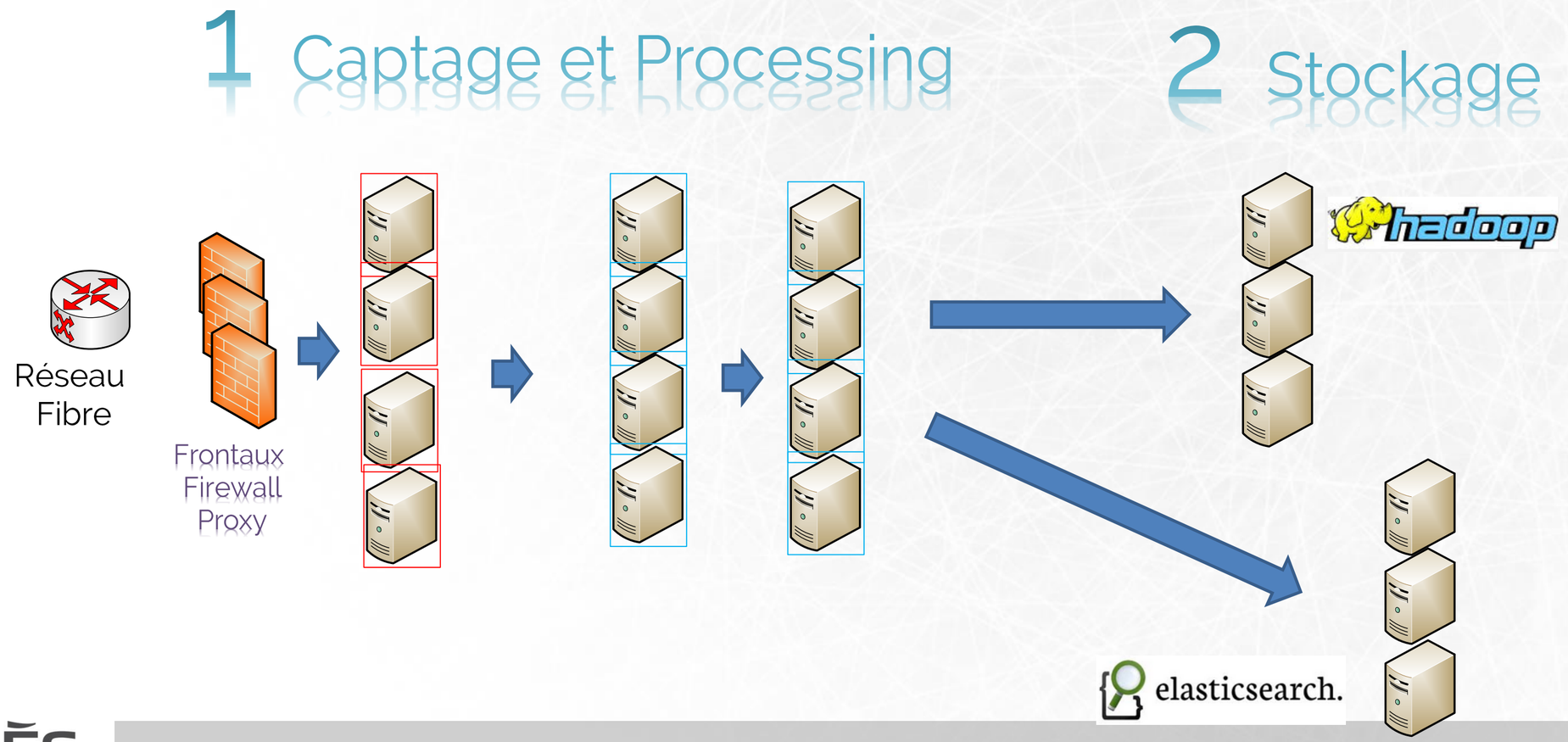
480 Go / Jour

200 To/ an

800 To/ 4ans

5- Un use case pour bien comprendre !!

Architecture Finale de processing Big Data



5- Use case N°2 : Prédire les Clients ?

- ✓ Détecter les clients qui vont signer
- ✓ Cibler les clients qui vont partir
- ✓ Adapter une action marketing pour les garder

Comment Faire avec le Big Data ?

Le futur du marketing : Deep Learning

Canaux Directs
Magasins…

Données internes
Rentabilités/ERP

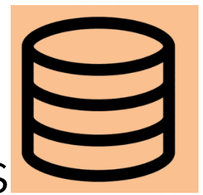
LesFurets.com
Comparez futé

assurland.com
le comparateur d'assurances

Comparateurs

Plateforme
Hadoop
Et outils de
prédiction

Site internet
Outils CRM
Outils interne



Données Internes
Devis/commandes

Données internes
Commerciales (CRM)

5- Use case N°2 : Machine ou Deep Learning C'est quoi ?

- ✓ Intelligence artificielle
- ✓ Algorithmes d'apprentissage (réseaux de neurones…)
- ✓ Connaissance du passé et du présent pour prédire l'avenir

Les gang des GAFA (Google, Apple, Facebook, Amazon)

En octobre 2015, le programme alphaGo ayant appris à jouer au jeu de go par la méthode du *deep learning* a battu par 5 parties à 0 le champion européen Fan Hui[3].
En mars 2016, le même programme a battu le champion du monde Lee Sedol 4 parties à 1

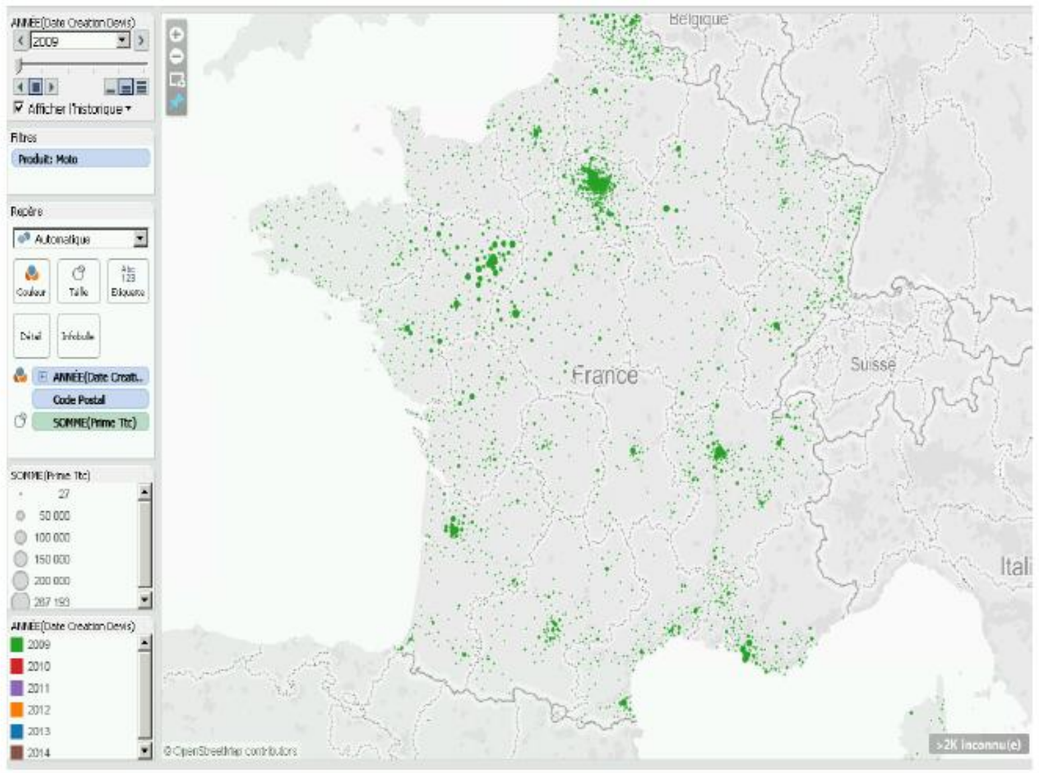
5- Use Case avec le big Data

- Connaissance du client
  - Affinité avec les canaux de communications
  - L'appétence par rapport aux offres produits

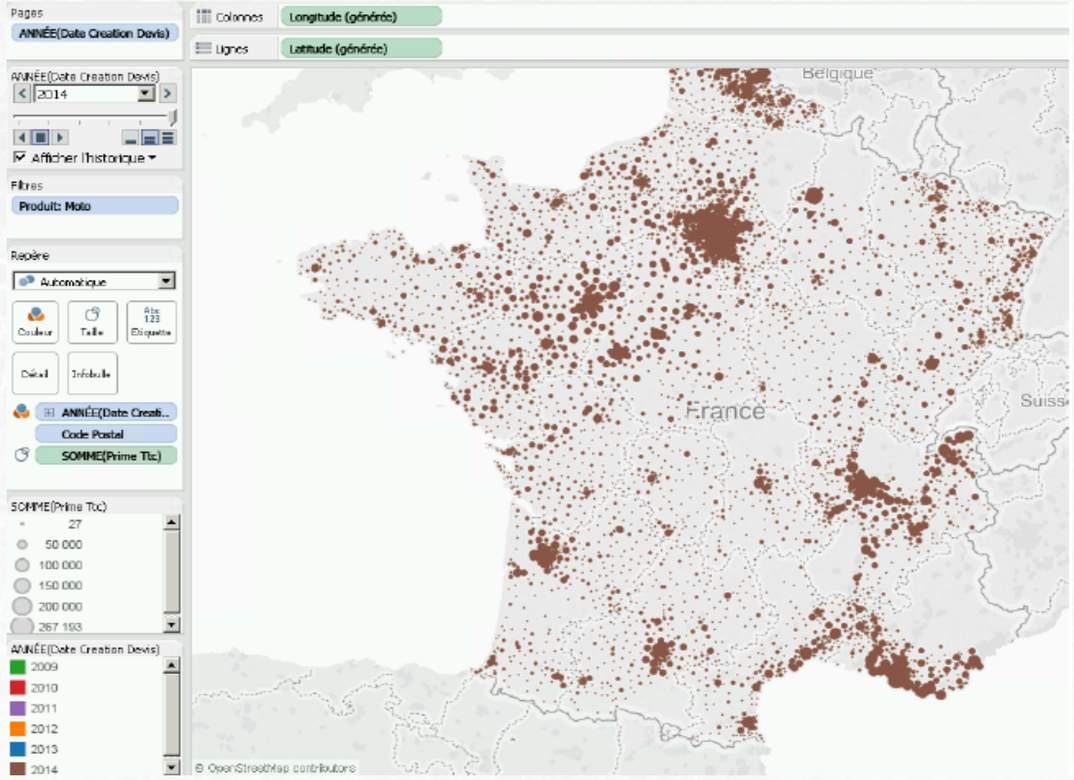
- Marketing et conception de produits nouveaux
- Analyse de la Fraude (Prédictif..)
- Scoring du churn
- Analyse des comportements du client en temps réel (Achalandage des magasins)

Avant : Analyse de la clientèle

2009



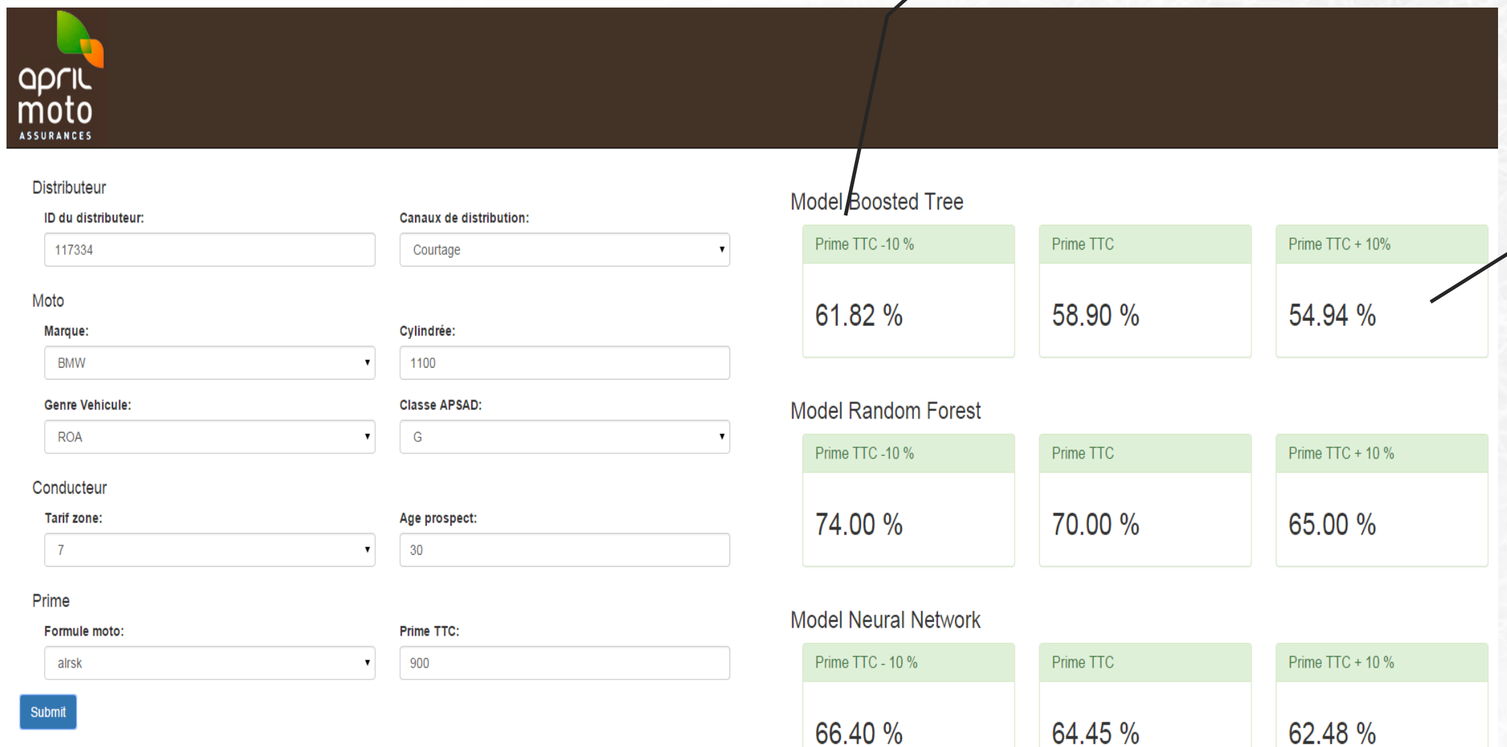
2014



Après !!!!

Probabilité de souscrire avec une baisse de tarif de 10%

Probabilité de souscrire avec une Hausse de tarif de 10%



Modèle d'apprentissage (Boosted Tree)

Modèle d'apprentissage (Arbres Aléatoires)

Modèle d'apprentissage (Réseau de neurones)

5- Use Case futuriste : L'assureur connecté

- Les objets connectés
  - Pay As You Drive
  - BodySelf (Montre Connectée)

- Conception de produit éphémères

  Assurance Vie One Click

  Assurance auto One Travel

- Réseau Social d'entraide (Startup Néolink)

Plan

1- Introduction
2 - Le Big Data c'est quoi ? Mythe ou réalité ?
3 - L'histoire et la genèse du Big Data
4 – Les applications majeures du Big data
5 - Les uses cases dans le marketing (Ciblage client et calcul du Churn)
6 – L'éthique dans tout cela ? La France protectionniste ?

6- l’ UE et les données personnelles

- Des règles communes pour l'UE ont dès lors été mises en place afin de garantir que vos données personnelles puissent bénéficier d'un niveau élevé de protection dans tous les pays de l'UE.

- Vous avez le droit de porter plainte et d'obtenir des mesures réparatoires si vos données sont utilisées à mauvais escient au sein de l'UE.

- La directive sur la protection des données de l'UE prévoit également des règles spécifiques pour le transfert de données personnelles à l'extérieur de l'UE afin de garantir la meilleure protection possible pour vos données transmises à l'étranger.

6- l’ UE et les données personnelles

Des données à emporter !

Plus de transparence

Protection des mineurs

2018

Guichet unique

Sanction renforcée

Consécration du droit à l’oubli

6- l'éthique et les données personnelles

- L'Europe et la France trop protectionnistes ?
  - ✓ Réglementation très dure en Europe , vers une protection totale des données personnelles
  - ✓ CNIL , Loi Informatique et liberté en France

- US et pays émergents , « Open Bar »
  - ✓ Législation ouverte pour le business
  - ✓ Le consommateur et ses données sont financiarisées

- C'est peut être une force pour l'UE ? Finalement