

# Le Big Data au service d'un développement socio-économique

Tuteur: M. Olivier Bernard

**Projet Personnel en Humanités**

## **Le Big Data au service d'un développement socio-économique**

## **Le Big Data au service d'un développement socio-économique**

### **Remerciements**

A mon tuteur Olivier Bernard qui m'a ouvert, grâce à son cours, au monde du machine learning.

A ma tutrice de stage, Marine Leclerc, qui a été mon mentor en Data Science durant 6 mois

A l'équipe PPH de l'INSA qui permet aux élèves ingénieurs d'explorer de nouveaux horizons à travers la réalisation de ce projet.

A ma famille qui m'a toujours soutenu dans les projets que j'ai entrepris

## Le Big Data au service d'un développement socio-économique

# Sommaire

|   |    |
|---|----|
| 1. Introduction.....  | 6  |
| 2. Qu'est ce que le big data.....   | 8  |
| 2.1 Le Big Data.....  | 8  |
| 2.2 Un monde interconnecté.....   | 9  |
| 3. Le Big Data pour le développement.....                                   | 11 |
| 3.1 Une expérience de pensée.....   | 11 |
| 3.2 Des données au service du développement.....                            | 12 |
| 3.3 Un Big Data pour le développement.....                                  | 13 |
| 4. Les technologies pour le Big Data.....                                   | 15 |
| 4.1 L'intelligence artificielle.....  | 15 |
| 4.2 Caractérisation d'une donnée.....                                       | 15 |
| 5. Un Projet concret.....   | 17 |
| 5.1 Le projet Jakarta smart city pour améliorer les transports publics..... | 17 |
| 5.2 L'analyse des données.....  | 17 |
| 5.3 Les résultats.....  | 18 |
| 6. Les Limites liées à la donnée.....                                       | 20 |
| 6.1 La vie privée.....  | 20 |
| 6.2 L'accès et le partage.....  | 20 |
| 6.3 La difficulté de l'analyse.....   | 21 |
| 6.4 Détecter les anomalies.....   | 22 |
| 7. Conclusions.....   | 24 |
| Bibliographie   |    |

## **Le Big Data au service d'un développement socio-économique**

## Le Big Data au service d'un développement socio-économique

### 1. INTRODUCTION

Le 21ème siècle est marqué par la prolifération des données dans tous les secteurs de nos vies et de la société. Avec le développement de nouvelles technologies d'information et de télécommunications, nous assistons à un réel déluge de la donnée dans nos vies de tous les jours. Si l'Homme dans l'histoire a toujours gardé des données sous type de format en commençant par les gravures sur des pierres, puis ensuite l'écriture et maintenant le numérique, aujourd'hui c'est la vitesse et la quantité à laquelle ces données sont générées qui fait rebattre toutes les cartes. Chaque jour 2,5 quintillions ( $10^{30}$ )<sup>1</sup> octets sont générés. Durant ces deux dernières années, 90% de données existantes ont été générées et cela n'est pas prêt de s'arrêter.

Bien que ces données soient stockées sous un format binaire à l'aide de 0 et de 1, elles peuvent représenter diverses choses. D'une image sur Facebook à une voix, en passant par la géolocalisation téléphonique des citoyens d'une ville ou encore des données des compteurs électriques et le nombre de messages transitant par une antenne relais.

Cette prolifération des données a permis le développement de nombreuses entreprises qui ont su tirer parti de ces données pour fournir à leurs clients un service plus ciblé qui correspond à leurs besoins. Le plus célèbre d'entre eux est bien sûr le géant du web Google qui juste avec son moteur de recherche, collecte par seconde 80 000 données<sup>2</sup>. Ces dernières correspondent aux recherches effectuées sur le moteur qui, nous le verrons plus tard peuvent permettre de prendre une multitude de mesures et d'actions.

Si le Big Data est aussi important, c'est qu'il permet de faire ressortir certaines informations et 'patterns' qui était invisible auparavant. D'après le dictionnaire en Collins<sup>3</sup>, un "pattern" est défini comme une façon régulière ou répétitive qu'un événement a à se répéter.

De plus, avec le développement de l'intelligence artificielle et des algorithmes de machine learning, qui existe depuis bien avant l'ère Big Data, la détection de ces "patterns" est beaucoup plus efficace car la performance de ces algorithmes est fortement liée avec la quantité de données dont ils disposent.

Figure 1 – Harvest data from Petabytes



source : <https://www.dtiers.com/big-data-2/>

## Le Big Data au service d'un développement socio-économique

Cela dit, bien que le Big Data ait permis à de nombreuses entreprises de faire fortune, il est peu commun d'entendre parler de Big Data pour le développement dans les pays pauvres, car si les pays du Sud rattrapent peu à peu leur retard sur le digital face au Pays du Nord il n'échappe pas au développement du Big Data. Lorsque nous parlons des pays du Sud nous parlons des pays sous-développés et/ou en voie de développement. À l'inverse, nous qualifions de pays du Nord, les pays où les populations ont accès à l'ensemble des besoins vitaux tels que l'éducation, la santé, l'eau potable ou encore la nourriture. Ces pays ont su rattraper leur retard sur le digital ce qui a provoqué des transformations profondes dans leurs sociétés. Le nouveau paradigme qu'amène le Big Data dans ces pays propose de nouvelles opportunités de développement qui ont longtemps été inenvisageables auparavant. La technologie permet aux preneurs de décisions de faire état du développement, d'améliorer la protection sociale et comprendre quels aspects de la législation ont besoin d'ajustement pour bénéficier au mieux les citoyens.

Ce sont ces nouvelles possibilités de développement qu'apporte le Big Data que nous nous proposons d'étudier dans ce travail de recherche. Nous nous concentrerons tout d'abord sur ce que représente les Big Data ensuite nous nous proposerons d'étudier dans quelle mesure il permet de contribuer au développement et les défis qui y sont associés. Ensuite nous présenterons un cas d'application pour enfin étudier les limites de ces technologies.

Figure 2 – What is Big Data for development



source: <https://www.slideshare.net/unglobalpulse/big-data-for-development-a-primer>

## Le Big Data au service d'un développement socio-économique

### 2. QU'EST CE QUE LE BIG DATA

#### 2.1 Le Big Data

Alors que nos générations précédentes possédaient peu de données sous format analogique et disponible sous peu de formes, nous assistons aujourd'hui à un déluge de données générées. Ce qui change par rapport au passé est la rapidité à laquelle nous avons accès à la donnée. Ensuite, la nature des informations change considérablement avec le développement qu'ont connu les réseaux sociaux et la multitude de services offerts par les téléphones et smartphones. Avec cette révolution, les données sont générées de manière autonome alors que les personnes agissent à leur gré. À tout moment dans le temps, ces données générées sont accessibles, ce qui crée de nouvelles opportunités pour superviser et suivre le comportement de différentes communautés.

Si la révolution du Big Data a pris son essor dans les pays industrialisés, elle a aussi lieu dans les pays en voie de développement et s'intensifie chaque jour. Elle se fait tout d'abord au travers du téléphone. Le taux de pénétration téléphonique a connu une importante croissance dans ces zones du monde. Entre 2011 et 2015, le nombre de téléphone est passé de 500 million à 850 million<sup>4</sup> alors que le taux de pénétration téléphonique n'atteignait même pas les 10% en 2002. On pense plus précisément aux applications qui en sont faites. Dans certains pays, le téléphone est devenu un substitut aux faibles infrastructures de télécommunications et de transport, mais aussi aux systèmes financiers et de banque.

Le téléphone est devenu plus qu'un outil de communication et sert aussi d'outil pour transférer de l'argent, un compte en banque, un outil de facturation ou encore un outil de sauvegarde de données médicales. En Afrique de l'Est, le service M-PESA<sup>5</sup> permet de faire des paiements mobiles aux banque et à d'autres individus ou SoukTel<sup>6</sup> au Moyen-Orient qui à tout d'abord commencer en tant que plateforme qui mettait en contact par SMS des chercheurs d'emploi avec des employeurs et qui maintenant développe la même offre sur le web.

L'utilisation des réseaux sociaux comme Facebook ou Twitter est aussi très importante dans ces pays. Radio France International rapportait en 2017<sup>6</sup> que le nombre de comptes Facebook sur le continent africain s'élevait à 146,6 millions et que l'on compte 100 000 nouveaux comptes par mois dans un pays comme le Sénégal. Pouvoir suivre les tendances sur les réseaux sociaux peut fournir d'importantes informations sur les préoccupations et les tendances de comportement des utilisateurs en temps réel à une échelle local et peut s'avérer très utile pour les organismes de développement. De plus, ces mêmes organisations de développement proposent des programmes aux populations vulnérables sont une autre source de données temps réel, particulièrement dans les cas où un service de numérisation est impliqué.

Bien que l'explosion des données ne se fasse pas de la même manière dans les différentes zones du monde et encore moins à la même allure, le fossé entre les régions se réduit plus vite que nous ne le pensions. Dans les pays avec des institutions peu développés, la révolution de la donnée devient importante pour pallier le manque de quantité ou de fiabilité des données. Il semblerait donc que le Big Data soit un bien facile d'accès qui crée de nombreuses opportunités pour les pays en voie de développement. Mais la donnée ne serait rien sans la capacité de la générer en grande quantité et la volonté de vouloir les analyser.



## Le Big Data au service d'un développement socio-économique

### 2.2 Un monde connecté

À l'heure de la mondialisation et d'un monde interconnecté, les situations socio-économiques de certains pays peuvent devenir très instables et certaines communautés se retrouvent très vulnérables face à certaines crises économiques. Joseph Stiglitz disait «qu'autrefois: lorsque les États-Unis éternuent, le Mexique s'enrhume. Aujourd'hui lorsque les États-Unis éternuent, une grande partie du monde attrape la grippe»<sup>7</sup>. Cette phrase est représentative de la situation à laquelle peuvent faire face certains pays pauvres qui sont exposés à une importante vulnérabilité économique. À cela ajoutons les risques de sécheresse, l'inflation, le taux de chômage, etc.

L'origine de ces crises est donc climatique ou financière. Si nous revenons un peu en arrière, la crise du pétrole de 2007 et 2008 a été suivie par une importante récession. En 2011, le monde est encore passé par une crise avec une famine<sup>8</sup> dans la corne de l'Afrique (Éthiopie, Kenya, Somalie, Djibouti, Soudan) due à une sécheresse qui menaçait la vie de plus de 10 millions d'habitants. D'après l'OCDE, les chocs économiques de grande envergure risquent d'avoir plus d'impact à l'avenir<sup>9</sup>. Les conséquences de crises financières et pandémiques vont s'intensifier avec l'interconnectivité de l'économie mondiale et la vitesse à laquelle les populations se déplacent ainsi que les biens et les données sont échangés.

Pour plusieurs foyers dans les pays en voie de développement, la volatilité des prix de l'alimentaire représente le risque auquel ils sont le plus exposés.

Dans un monde où le système financier est interconnecté, il devient très difficile de renseigner l'impact local de manière immédiate et visible. En 2010, Fuentes Nieva Ricardo montrait dans «Shocks and Human Development»<sup>10</sup> qu'un impact socio-économique sur une communauté vulnérable pouvait avoir des conséquences sur plusieurs générations. Des enfants qui sont forcés d'être déscolarisés ne retrouvent jamais les chemins des bancs de l'école, ou encore une famille qui se retrouve dans l'obligation de vendre leurs biens ou migrer à cause de la peur de se retrouver dans une situation précaire. Des fœtus souffrant de malnutrition maternelle auront très probablement de séquelles une fois nés et pourront même en mourir.

Le temps que les informations renseignant sur ces problèmes remontent et alertent les preneurs de décision ou les journaux, il est souvent déjà trop tard où excessivement coûteux de répondre et proposer une solution efficace et durable. Les causes principales de ces catastrophes peuvent être multiples: une sécheresse, des inondations, une pandémie, un conflit armé, une pénurie de carburant ou un choc financier. Cela dit, même avec des outils de haut niveau il reste très difficile de distinguer quel groupe social sera plus ou moins affecté et dans quelle mesure.

Mais les gouvernements ne restent pas les bras croisés, et ont pris conscience de la gravité et des conséquences de ces catastrophes et réalisent aussi qu'il revient beaucoup moins cher de prévenir ou du moins de les contenir que de renverser une situation de crise. De nombreux organismes internationaux font office de lanceur d'alertes (UNICEF, WPF, PNUD)<sup>11</sup>.

L'information remontée par les sondages renseigne aussi de manière pertinente, mais ces données prennent du temps à être récoltées, traitées et vérifiées pour ensuite être publiées. Ces sondages sont trop difficiles à mettre en œuvre et coûteux pour pouvoir servir de solution. Cependant, ces données traditionnelles continueront de donner un aperçu relativement correct des situations de développement mais l'information digitale est une opportunité pour gagner en richesse et en pertinence d'information pour rendre compte d'indices de développement encore plus actuels.

## **Le Big Data au service d'un développement socio-économique**

Nous parlions précédemment du secteur privé qui a accueilli le Big Data à bras ouverts pour créer de la valeur. Nous entendons de plus en plus le discours prôner par ces grandes entreprises sur la valeur ajoutée de la donnée en temps réel qui l'or noir du 21ème siècle<sup>12</sup> et l'importance de la prise de décision fondée sur les données ("data-driven decision" en anglais)

La société civile commence à comprendre aussi l'importance du Big Data et accueille à bras ouverts la donnée massive en temps réel. Ce phénomène est mis en évidence par diverses méthodes comme le crowd-sourcing<sup>13</sup> qui ont pour but de réunir des communautés et des individus possédant une pensée commune via l'utilisation de téléphone ou d'autres plateformes comme internet, la radio etc. Dans plusieurs de ces situations, ces différents domaines constituent une nouvelle manière de commercer ou de faire du business.

Ces dernières années, les gouvernements ont compris le pouvoir du Big Data. Certains ont choisi de l'exploiter de manière plus conservatrice en imposant d'importantes mesures de contrôle alors que d'autres ont décidé de pousser les initiatives collectives avec des mouvements open data qui nous l'espérons pourra se servir de la puissance collective pour le bien commun.

Par exemple, les Nations Unies ont mis un point, à l'initiative de leur secrétaire générale, le laboratoire Global Pulse<sup>14</sup> avec pour but de développer une nouvelle approche du "social impact monitoring" qui est un terme signifiant le suivi des conditions socio-économiques d'une masse d'individus. L'objectif est d'analyser et de comprendre le comportement d'une population en utilisant de nouvelles sources de données comme nous l'avons présenté précédemment.

Au-delà de la donnée brute et de la volonté des institutions de les exploiter, les notions de "capacité" et de "moyen", sont indispensables pour pouvoir faire parler les données pour contribuer à un développement.

## Le Big Data au service d'un développement socio-économique

### 3. LE BIG DATA POUR LE DÉVELOPPEMENT

Le Big Data est un terme souvent utilisé pour décrire le volume massive de données, structuré ou non, si important, qu'il en est très difficile d'en extraire des informations à l'aide des bases de données ou méthodes informatiques traditionnelles. Les caractéristiques par lesquelles se distingue le Big Data sont souvent notés comme les 3 V<sup>15</sup>: Volume, Vitesse, Variétés.

Pour commencer, le volume. La donnée vient de plusieurs sources. De capteurs, de réseaux sociaux d'appareils connectés. Ces appareils génèrent une quantité abondante de données. Ensuite vient la vitesse à laquelle sont générées ces données. Les vidéos postées sur Internet, les signaux GPS émis, l'historique des achats effectués sur Amazon pour en nommer quelques-uns. Enfin la variété, les données peuvent être stockées sous forme d'une image, d'un son, d'un fichier texte, d'un tableau de données. Chaque type de données doit subir un traitement et une analyse adaptés au format sous lequel elle se trouve.

Maintenant que nous avons défini le Big Data, nous allons pouvoir présenter dans quelle mesure le Big Data peut être utilisé dans un contexte de développement. Pour cela, nous allons réaliser une expérience de pensée avant de présenter des cas d'utilisation récents et concrets. Par expérience de pensée nous entendons la présentation d'un scénario fictif mais fortement plausible qui est conforme à la réalité.

#### 3.1 Une expérience de pensée<sup>16</sup>

Imaginons un foyer de 6 personnes, 2 parents et 4 enfants vivant à la périphérie d'une ville de taille moyenne à quelques heures de la capitale dans un pays en voie de développement. Le membre principal de foyer (source de revenus principaux) est un mécanicien possédant un petit garage. Sa femme cultive des légumes et élève des chèvres sur la portion de terre qu'ils possèdent qu'elle vend au marché en ville ainsi que de la couture et des draps. Leurs 4 enfants ont entre 6 et 18 ans. Le problème de la famille est le suivant. Elle a grandement souffert de la flambée des prix des produits de base (nourriture, essence, etc) ces derniers mois et risque de tomber dans la précarité. Pour faire face à la crise, plusieurs options se profilent.

Les parents pourraient commencer à réduire leurs dépenses en nourriture en trouvant des alternatives moins chères ou acheter des tanks de réserve(eau ou lait) ou tout simplement sauter des repas. Ils pourraient aussi commencer à s'approvisionner à un centre de distribution World Food Programme.

Pour réduire leurs autres dépenses, le père pourrait commencer à travailler plus tôt le matin pour pouvoir revenir moins tard le soir avant le coucher du soleil et économiser sa facture d'électricité. La mère pourrait travailler plus d'heures et partir au centre-ville plus souvent pour vendre plus de marchandises. De plus, diminuer son forfait de téléphone qui fonctionne avec un système de carte pré-payée. Au lieu de fonctionner par un gros incrément en début de mois elle pourrait faire de petit incrément pour gérer au mieux ses dépenses téléphoniques. Elle pourrait aussi retirer de l'argent de ses économies qu'elle aurait mise sur un système de mobile-banking.

Si la situation empirait, le père serait dans l'obligation de vendre des pièces de son garage ou quelques chèvres ou encore ne pas payer leur micro crédit pour leur maison. De plus, ils pourraient appeler des proches expatriés en Europe, pour une aide financière temporaire. Une option est aussi

## Le Big Data au service d'un développement socio-économique

de déscolarisé temporairement le plus jeune de leurs enfants pour économiser de l'argent sur les frais de scolarités. Si après toutes ces mesures drastiques, la situation ne s'améliore toujours pas, le plus jeune de leur enfant pourrait montrer des signes d'anémie en raison de la faible alimentation et des coupures du budget alimentaire. Les parents commenceraient alors à consulter les lignes téléphoniques gratuites de santé alors que le plus grand des enfants ferait des recherches sur internet au cyber café du quartier pour comprendre les symptômes de son frère et se plaindre sur son mur Facebook de la situation socio-économique de sa famille comme bien d'autres jeunes de la région qui souffrirait de cette crise. C'est alors que l'aide local prendrait conscience du problème pour commencer à prendre des mesures. De même que les journalistes qui commencerait à reporter la situation dans la presse et dans des organismes internationaux.

### 3.2 Des données au service du développement

Un choc d'une telle ampleur peut toucher des dizaines comme des milliers de foyer et d'individus. On peut imaginer que de nombreuses familles adopteraient le même comportement que la famille que nous avons décrits dans notre expérience de pensée.

Tout au long du scénario, le comportement que de nombreuses familles pourraient adoptées de manière simultanée, fournirait de nombreux indicateurs et seraient générateurs de nombreuses données. Par exemple:

- L'opérateur mobile a pu se rendre compte d'une baisse de la quantité d'abonnés qui ont réduit le coût de leur forfait au début du mois. Les données pourraient aussi montrer une baisse importante dans le nombre d'appels passé et une augmentation de l'échange de SMS dont l'envoi est généralement moins coûteux.
- Les services de mobile banking auraient noté que leurs clients retireraient leurs argents de leurs comptes virtuel. De plus, quelques semaines après que la situation ait commencé à empirer, un défaut de paiement dans les prêts aurait pu mettre la puce à l'oreille aux observateurs.
- Les lignes gratuites d'assistance santé téléphonique auraient enregistré de nombreux appels reportant les symptômes d'une malnutrition.
- D'autres sources comme les réseaux sociaux auraient fait été de la situation. Par exemple sur le nombre de Tweets en lien avec la difficulté de se procurer de la nourriture en raison des prix.
- En raison du manque de moyen les personnes utiliseraient moins leurs voitures ou camions pour se déplacer. Ce phénomène pourrait être représenté via l'analyse d'images satellite.
- Les centres de distributions de nourriture serviraient plus de nourriture qu'habituellement et les services d'aide à l'enfant (UNICEF), qui suivent de très près le taux absentéisme dans les écoles se rendraient vite compte d'une baisse.

L'exemple que nous avons peint précédemment montre l'analyse qu'il est possible de tirer de la donnée en temps réel et dans quelle mesure il peut permettre de prévenir certaines situations de crises.

## Le Big Data au service d'un développement socio-économique

Maintenant que nous avons présenté les possibilités proposées par les données dans le secteur du développement, nous allons définir plus en détails quelles sont les caractéristiques, les sources et les catégories de Big Data qui peuvent être utiles pour le développement.

### 3.3 Un Big Data pour le développement

Il faut bien comprendre que le Big Data pour le développement diffère grandement des termes traditionnels comme “la donnée pour le développement” (données de sondages ou statistiques officielles) ou encore ce que les médias et le secteur privé appellent le “Big Data”<sup>17</sup>. Par exemple, les données issues de la microfinance (le nombre de clients, le montant et les types de prêts, les défauts de paiement) sont au croisement entre la donnée traditionnelle de développement et le Big Data pour le développement en raison de la nature de l'information qui peut s'avérer utile pour les experts des questions de développement. Avec la prolifération de la gestion d'argent autour du téléphone et des plateformes en lignes pour recevoir des micros crédits, une quantité importante de données de microfinance peut être analysée en temps réel pour donner état des conditions de vies d'une population. De ce fait, ce type de donnée peut être considéré comme Big Data pour le développement.

Un autre exemple de Big Data pour le développement peut être présenté au travers du prisme des Tweets, des données mobiles, des requêtes en ligne. Ces données s'ancrent tout à fait dans ce que la foule appelle le Big Data (une grande quantité de données générées de manière passives par des utilisateurs à une haute fréquence). Cela dit, bien que ce type de données n'est pas utilisé traditionnellement pour le développement, elles peuvent devenir de précieux indicateurs des conditions de vie d'un groupe de personnes et qui donc peut être inclus dans le Big Data pour le développement.

Les sources de Big Data pour le développement partagent souvent plusieurs des caractéristiques que nous allons citer ci-dessous.

- Générées numériquement - Les données sont générées numériquement et non manuellement (sondages) et sont stockées sous format binaire et peuvent être manipulées directement par des ordinateurs.
- Produites de manières passives – Ces données sont produites par des produits de la vie de tous les jours (téléphone, services digitaux en ligne)
- Collectées de manière automatique – Un système en aval de la source extrait et sauvegarde les données utiles générées de manière autonome.
- Information de temps et d'espace liée à la donnée – la donnée est toujours taguée d'une information spatio-temporelle (par exemple les téléphones sont géo-localisable et la durée des appels peut être enregistrée).
- Analysées en continu – l'information peut être analyser en continu pour répondre au mieux à une situations

Il est important de comprendre que dans le cas du développement international, le terme “temps réel” (“real-time” en anglais) ne signifie pas toujours instantanée, mais plutôt temps réel dans le sens où l'information produite est rendue accessible dans un temps relativement court sur

## Le Big Data au service d'un développement socio-économique

laquelle la donnée gardera sa pertinence. On parle de pertinence, car une fois l'information reçue, une action ou une décision peut être prise de manière efficace. Nous retenons par-dessus tout que c'est la dimension temporelle liée à la prise de décision qui est intrinsèquement liée à la notion de temps réel de la donnée. On ajoutera aussi que la nature temps réel d'une donnée est liée à l'analyse qui en faite et par extension à l'action qui est prise.

Attardons-nous maintenant sur la granularité de la donnée. Dans le Big Data pour le développement, les données récoltées aux plus petites échelles ne sont pas nécessairement les plus utiles. Des données agrégées à l'échelle d'une communauté peuvent être dans certains cas préférables plutôt que des données individuelles (à l'échelle de la personne ou d'un foyer) et mieux protéger les vies privées. Pour ce qui concerne la dimension temporelle, il est compliqué de définir des normes qui pourront rendre compte de la pertinence de la donnée puisque la vitesse à laquelle elles sont générées varie très fortement. La notion de donnée temps réel restera toujours une notion éphémère et tout ou tard, la donnée se transformera en donnée obsolète. On peut penser à l'historique de pluie d'un lieu précis, un historique de prix ou encore des données statistiques d'organismes officiels comme le PIB.

Nous pouvons identifier plusieurs sources de données qui peuvent s'avérer utiles pour contribuer au développement. Ces dernières ont été recensées par le laboratoire Global Pulse des Nations Unies<sup>18</sup>.

- Données générées involontairement – Les données récoltées des services numériques comme les smart phones, les achats en ligne, les recherches internet etc.

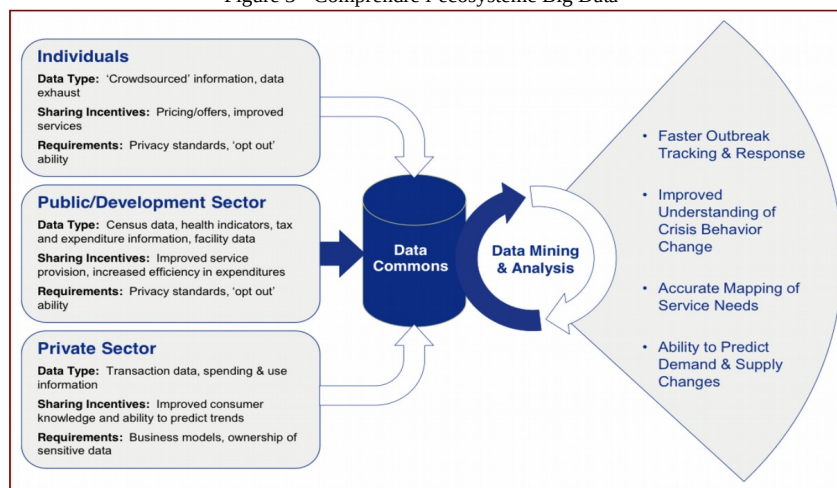
- L'information d'internet: À ne pas différencier avec la source précédente, ces données correspondent à tout le contenu des pages web, des réseaux sociaux, des articles de presse, etc.

- Capteurs : On pense aux capteurs géologiques, aux images satellites, aux émissions de lumières qui rendent compte de l'activité et du changement de comportement des individus

- Données soumises volontairement: Enfin nous pouvons retenir les informations produites par les citoyens et les organismes officiels au travers de sondage en ligne ou téléphonique. Ces informations sont indispensables pour pouvoir vérifier la pertinence des données générées par les sources ci-dessus.

Pour conclure sur les données sur le développement voilà un schéma(anglais) publié par weforum en 2012 qui présente une autre perspective sur la manière dont la donnée peut être générée, mis à disposition et dans quelle mesure une action peut être prise.

Figure 3 - Comprendre l'écosystème Big Data



Source : Le nouvel écosystème big data, Illustré par le forum économique mondial, Source: WEF White Paper, "Big Data, Big Impact: New Possibilities for International Development" [http://www3.weforum.org/docs/WEF\\_TC\\_MFS\\_BigDataBigImpact\\_Briefing\\_2012.pdf](http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf)



## Le Big Data au service d'un développement socio-économique

### 4. LES TECHNOLOGIES POUR LE BIG DATA

Les technologies pour le big data ont gagné en moyen, en présence sur la place publique, en finance, et en performance ces cinq dernières années. A l'origine utilisées dans le secteur de l'électronique, la biologie, de la médecine, les technologies du Big Data font référence aux outils et aux méthodes qui ont pour but de transformer des quantités massives de données brutes en données lisibles qui permettent de traduire un phénomène, une tendance, des corrélations, pour faire simple en une information compréhensible pour l'humain pour qu'il puisse prendre une décision en conséquence.

#### 4.1 L'intelligence artificielle

Le Big Data repose sur des algorithmes puissants (de machine learning par exemple) qui sont capables de repérer des "patterns" sur de larges jeux de données, mais aussi permettent de développer des outils de visualisations très avancées qui sont indispensables à la compréhension de l'information.

Ces outils et algorithmes permettent d'analyser la donnée, mais aussi de prédire les comportements futurs du système dont l'information de sortie est renseignée par les données récoltées. Pour cela, le système va se baser sur les données initialement récoltées et va pouvoir établir un modèle qui permettra de déduire les issues futures en fonctions des données d'entrée futures. L'établissement de ce modèle est appelé entraînement. Une fois cette étape effectuée, ces algorithmes permettent de prendre des décisions qui permettent de détecter des incohérences ou des anomalies sous forme d'une différence, d'une déviation, d'un écart entre les nouvelles données et les données précédemment traitées.

Pour pouvoir déceler des "patterns" et des tendances dans les données, issues de plusieurs sources d'informations, il faut tout d'abord définir un cadre, une démarche dans la chaîne de traitement de données.

#### 4.2 Caractérisation d'une donnée

Quatre segments d'information sont importants dans la caractérisation de la donnée<sup>19</sup>:

- QUOI: Quel type d'information contient la donnée
- QUI: a rapporté la donnée
- COMMENT: Par quel canal la donnée a été remontée
- Où et QUAND: La dimension spatio-temporelle de la donnée, à quelle échelle est-elle récoltée et quel est intervalle de temps entre lequel la donnée a été générée et la donnée a été récoltée.

Une fois la donnée taguée, elle doit ensuite être traitée avant d'être analysée. Cette étape peut inclure:

## Le Big Data au service d'un développement socio-économique

- Le filtrage: Récupérer uniquement l'information utile en faisant abstraction des valeurs aberrantes
- Agrégation: il faut souvent passer par un agrégat de données avant de pouvoir les analyser (moyenne ou somme).
- Catégorisation: Certaines données brutes doivent être transformées pour être ensuite des indicateurs pertinents.

Une fois la donnée prête à être analysée, nous pouvons appliquer des algorithmes de machine learning pour la modéliser. Les algorithmes de machine learning pourraient être un livre en soi, c'est pour cela que nous n'allons pas les décrire en détail. Une des caractéristiques de ces algorithmes puissants, et qu'ils sont capables de s'adapter à de nouveaux flux de données. L'adaptation en continu est une composante essentielle dans la mesure où ces algorithmes s'améliorent au fur et à mesure que la quantité de données augmente.

L'extraction intensive de données socio-économiques est connu sous le nom de "reality mining"<sup>20</sup>.

Il permet de montrer au grand jour des processus de traitement et des interactions dans les données qui étaient encore inconnus. Cette extraction peut se réaliser de trois manières différentes:

- Analyse continue des flux de données: Utilisation d'outils pour récolter des données en ligne et les analyser.
- Récole des données semi-structurées: comme les article de presse, les avis sur les sites qui rendent compte des tendances et des perceptions des populations
- Les corrélations en temps réel avec les flux de données haut débit, mais un accès lent pour garder de grand historique. Cette dernière méthode consiste à faire une analyse rapide, mais sur de grand jeu de données qui feront mieux état de la situation.

Ces trois techniques peuvent s'avérer utile pour le Big Data à différent niveaux mais dépendra fortement de la donnée disponible.

Pour terminer sur cette partie, il ne faut pas oublier l'importance de la visualisation dans le Big Data pour le développement qui ouvre de nouvelles perspectives d'analyse. De plus, elle permet de dénoter de nouvelles découvertes qui sont d'habitude très difficile à déceler.



## Le Big Data au service d'un développement socio-économique

### 5. Un projet concret

Maintenant que nous avons couvert les possibilités du Big Data pour le développement, nous allons observer les différents projets d'application de Big Data pour le développement. Ils ont été entrepris par des acteurs du secteur public ou privé. Les cas que nous allons présenter sont des cas concrets qui illustrent des possibilités parmi tant d'autres d'application de ces technologies.

#### 5.1 Le projet Jakarta smart city<sup>20</sup> pour améliorer les transports publics.

Un projet qui est un parfait exemple de l'usage du Big Data pour le développement, est le projet mené par Jakarta Smart City en collaboration avec le Laboratoire Global Pulse. Le premier est un programme d'initiative lancé par la ville pour intégrer la ville dans une démarche de smart city. Quant au second, il a déjà été présenté plus haut dans notre travail. L'objectif de ce projet était d'améliorer la planification du réseau de Bus TransJakarta en utilisant les données de ce dernier. Il fallait en partie pouvoir cartographier les goulets d'étranglement pour identifier la mise en place de nouveaux itinéraires. De plus, ce projet avait aussi pour but d'explorer les données en temps réel du temps d'attentes des passagers pour pouvoir améliorer la répartition des bus sur les différentes lignes.

La ville de Jakarta est très connue pour ses embouteillages interminables ce qui prolonge les temps des trajets des habitants de la ville en début et en fin de journée ce qui complique la programmation des transports publics. C'est avec une volonté d'améliorer le service de transport que ce projet a été porté.

Établie en 2014, la plateforme Jakarta Smart City collecte des données issues du gouvernement, de parties tierces ainsi que des données issues du crowdsourcing. La plateforme inclut aussi des données sur le réseau de bus TransJakarta. Chaque année, plus de 120 millions de passagers empruntent ces bus. Les défis adressés dans cette étude étaient la fréquence du service, les temps de trajets ainsi que les temps d'attentes. Cette démarche a été adoptée pour acquérir un aperçu du comportement des passagers sur la donnée du trajet et le temps d'attente.

La flotte de bus de TransJakarta possède des GPS qui permettent de les localiser en temps réel toutes les cinq secondes. Les données remontées sont la géolocalisation, la vitesse et l'heure à laquelle elle est renvoyée. Ce sont en premier lieu ces données-là qui ont été utilisées pour identifier les lignes les plus lentes ainsi que les zones de la ville où la probabilité d'occurrence d'un embouteillage est la plus grande.

Durant le projet les données enregistrées par les bornes de validation des tickets ont aussi été utilisées pour renseigner sur le comportement des passagers. Ces données ont majoritairement servi à l'identification des lieux de départ et de destinations des passagers mais aussi le type d'abonnement et les heures de trajets. En faisant la différence entre l'heure d'achat du ticket, codée sur la bande du ticket et l'heure de validation du ticket sur la borne dans le bus, il est possible de retrouver le temps d'attente d'un passager. C'est de cette manière que le temps d'attente moyen d'un passager a dû être déterminé, sans avoir eu accès aux informations personnelles de ce dernier.

Les données ont été récoltées sur une période de deux mois sur 12 lignes différentes circulants dans toute la ville de Jakarta.

## Le Big Data au service d'un développement socio-économique

### 5.2 L'analyse des données

Le projet s'est concentré sur deux aspects. Le premier repose sur les données des GPS des bus dans le but de pouvoir cartographier les zones sujettes à des problèmes de circulation. Le second est associé aux données générées par l'utilisation du transport par les passagers. Ou le moment où des passagers commencent un trajet. Pour mieux comprendre l'utilisation des bus de voyageurs, le projet a exploré le tap-in (validation des tickets aux bornes) de mai et juin 2016 et a sélectionné une période d'une semaine pour analyser les tendances hebdomadaires. L'analyse comprenait 12 lignes et 422 694 passagers uniques, dont 202 933 des passagers (48%) ont été inclus dans l'étude origine-destination.

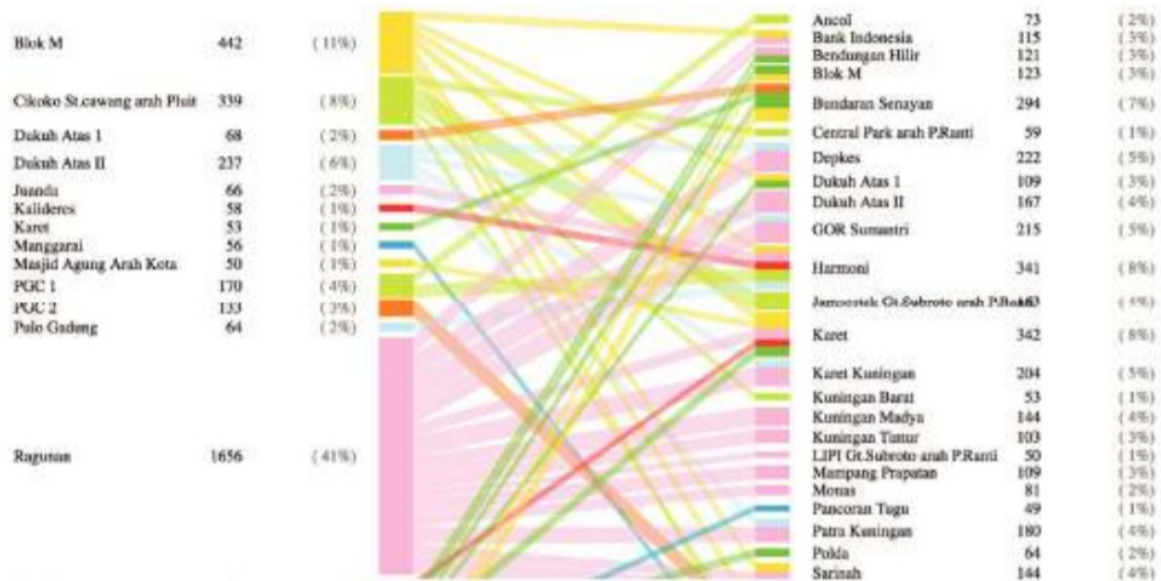
Pour déduire les statistiques origine-destination, le projet a développé deux approches différentes. Premièrement, sur la base des transactions de passagers durant la journée, quant à l'origine-destination, elle a été dérivée de deux étapes consécutives : une à l'origine et la seconde à destination. Les lieux de résidences et de travail des passagers ont été déterminés avec les lieux de validation de leurs cartes d'abonnement le matin et le soir.

### 5.3 Les résultats

Sur la base de ces deux méthodes, plusieurs éléments ont été déterminés.

- Le nombre d'heures entre deux voyages consécutifs
- Identifier les pics dans le trafic
- Les stations les plus desservies
- la paire origine-destination la plus fréquente à différents niveaux: stations, quartier, de la ville

Figure 4 - Visualisation des points de départ et d'origine des passagers utilisant le réseau TransJakarta illustrant les problèmes de congestions



Source: <https://beta.unglobalpulse.org/document/using-big-data-analytics-for-improved-public-transport/>

Figure 5 - cartographie des lignes du bus problématiques avec les zones de congestions

## Le Big Data au service d'un développement socio-économique



Source: <https://beta.unglobalpulse.org/document/using-big-data-analytics-for-improved-public-transport/>

Une fois ces caractéristiques du réseau identifiées, les résultats de l'étude ont été partagés avec le réseau TransJakarta pour permettre d'améliorer le dispatching des bus et actualiser certains itinéraires. De plus, l'étude a permis de découvrir certaines situations comme les raisons de certains voyageurs qui n'effectuaient le voyage que dans un sens le matin sans faire le retour le soir.

Ce projet est un exemple d'application du Big Data pour le développement qui peut servir à améliorer la fluidité du trafic d'une ville et dans quelle manière des structures publiques de transport peuvent générer des données avec beaucoup de valeurs qui peuvent grandement améliorer la condition des citoyens d'une ville.

Les possibilités paraissent donc infinies et d'autres vont voir le jour au fur et à mesure que les technologies se développent et que de plus en plus de données sont générées. Cela dit, le Big Data pour le développement fait face à de nombreux défis qui ne seront pas facilement surmontables. Certains de ces défis sont relatifs à la donnée elle-même (accès, partage, analyse) et à la thématique de la vie privée, tandis que d'autres sont liés à la difficulté d'analyse.

## **Le Big Data au service d'un développement socio-économique**

### **6. LES DIFFICULTÉS LIÉES A LA DONNÉE**

#### **6.1 La Vie privée**

Quand il s'agit des données, le sujet de la vie privée est peut être le plus sensible et le plus compliqué à aborder.

La vie privée est définie comme le droit aux individus de définir et de contrôler quelles informations les concernant directement sont en mesure d'être divulguées. La vie privée peut aussi être définie dans un sens plus large en englobant les entreprises qui souhaitent préserver leurs compétitivités et clients ou encore les états qui souhaite préserver leurs souverainetés et leurs citoyens face aux puissances étrangères. Dans les deux cas présentés, la vie privée est une préoccupation majeure que l'on retrouve dans différents cas d'usage pour quiconque s'intéresse au Big Data pour le développement. Du point de vue de l'acquisition, de la sauvegarde et de la conservation.

La vie privée est un droit humain fondamental qui a autant une valeur intrinsèque qu'instrumentale.

En 2010 déjà, Dirk Helbing et Stefano Baliotti écrivait dans « From Social Data Mining to Forecasting Socio-Economic Crisis » que « Une société moderne devait comprendre l'importance du privée pour pouvoir se développer ». Sans vie privée, sécurité diversité, et innovation, nos libertés les plus fondamentales sont menacées. De plus, ces risques touchent aussi les individus qui n'ont rien à se reprocher ou à cacher. Concernant les grandes structures, il n'est pas nécessaire de rappeler à quel point les informations manipulées sont sensibles et pourquoi la notion de privée est importante.

Attardons-nous maintenant sur le concept de vie privée de l'individu. Il arrive souvent que les utilisateurs de services digitaux générateurs de données n'ont pas conscience de ces derniers et quelle en sera l'usage effectué en aval. La majeure partie du temps, les utilisateurs cochent une case sur internet sans s'attarder sur l'utilisation qui est faite de leurs données. Il est aussi peu clair de la mesure dans laquelle les données seront analysées. Même si des efforts ont été entrepris dans certains pays comme la France avec la RGPD. Cela dit il y a toujours un manque de transparence pour des sites massivement visités comme Twitter ou Facebook.

La quantité d'information à l'échelle d'individu que possèdent les entreprises du numérique (GAFAM), les compagnies téléphoniques, les banques s'est rapidement avérée problématique. Par-dessus tout, la vie privée étant un pilier de la démocratie, il est important que les états et les organismes en charge prennent des mesures qui permettent de respecter les droits fondamentaux des citoyens et restent en alerte face à l'utilisation des données.

#### **6.2 L'accès et le partage**

Pour continuer sur les défis liés à la donnée, le second auquel fait face le Big Data pour le développement et l'accès et le partage.

Bien qu'une grande quantité de données présentent en accès libre sur les sites web open-data représentent un fort potentiel pour le Big-Data pour le développement, les données détenues par les entreprises ont beaucoup plus de valeurs. Un défi est d'aborder la réticence qu'ont les entreprises

## **Le Big Data au service d'un développement socio-économique**

privées et organismes à partager leurs données sur leurs clients et utilisateurs. Ces acteurs du Big Data ont peur que leurs compétitivité et réputation soient affectées. Ou encore la barrière légale qui n'autorise pas ces entreprises à partager ces informations ou encore une culture du secret fortement présente dans ce type de structure.

Ces défis sont aussi institutionnels et techniques. Nous le présentons plus tôt dans notre travail de recherche, mais l'état sous lequel est stockée la donnée détermine grandement sa capacité à être analysé. Certaines compagnies téléphoniques utilisent encore des supports analogiques pour sauvegarder ces données ce qui rend l'accès à la donnée d'autant plus difficile. Certaines compagnies téléphoniques stockent encore des données sur bande magnétiques.

Il faut donc être capable de signer des accords entre des partenaires privés et publics pour avoir accès à des données privée qui découlera d'arrangements légaux difficilement mis en place dans un premier temps. Ensuite, il faudra mettre en place des canaux de transfert pour que la donnée soit acheminée sur les serveurs et être analysée en continu et servir à l'entraînement des modèles. Ces derniers soulèvent tout de suite des problèmes d'inter-compatibilité entre les données et d'interopérabilité des systèmes numériques de télécommunications (protocole de communication, support physique de transfert et de sauvegarde).

Pour le Big Data pour le développement soit considéré comme tangible, il est indispensable de surmonter ces défis. Tout projet visant à contribuer au développement humain doit s'assurer que la vie privée des individus est préservée. Ces derniers se doivent de participer au débat public pour faire valoir leurs droits et s'approprier leurs données pour en comprendre la valeur. Toutes ces thématiques doivent nourrir le débat public sur la vie privée et la donnée pour permettre de proposer de nouvelles solutions et la mise en point de système plus respectueux des valeurs et des droits du numérique.

Cela dit, sans le soutien des institutions et des gouvernements, et par-dessus tout des acteurs privées pour partager leurs données, il sera impossible d'avancer. De nombreuses communautés émergent pour récolter des données et la anonymisés pour pouvoir contribuer. On pense à la communauté française Data for good qui œuvre sur de projets pour l'intérêt général comme l'étude des attentes des élections européennes de 2019 pour comprendre les attentes citoyens encore la promotion de plus de transparence dans les algorithmes donnant accès à l'information.

### **6.3 La difficulté de l'analyse**

Le dernier défi auquel nous allons faire face est l'accès à l'information. La plus grande difficulté quand il s'agit de Big Data pour le développement réside dans l'analyse et la question que tout spécialiste de la donnée chaque jour est : qu'est ce que me dit la donnée. Ce challenge possède plusieurs composantes. Le premier est la bonne identification du problème et qu'est ce qui est attendu de l'analyse. Ensuite comment bien interpréter la donnée et enfin détecter des anomalies dans la donnée.

La première étape dans une analyse de donnée et de vérifier la véracité de cette dernière. Des rapports non-vérifiés, une information falsifiée, des journalistes corrompues et tant d'autres de facteurs qui font que la donnée peut être fausse et en conséquence biaiser l'analyse d'un jeu de donnée massif. D'autres facteurs externes peuvent intervenir en jeu. Prenons l'exemple d'une situation où le flux de SMS envoyé serait utilisé pour mesurer la violence public. Les auteurs de ces actes chercheraient à limiter les signalements et donc le flux de SMS ne renseignerait plus sur la



## Le Big Data au service d'un développement socio-économique

localisation des téléphones mais seulement sur les téléphones que les auteurs ne peuvent pas supprimer, ce que nous appelons de “faux négatif” en statistique serait beaucoup plus importants dans les zones où il y aurait peu de violences alors qu’il y a simplement peu de trafic de SMS. Dans la majeure partie de ces cas, il y a une volonté d’altérer la perception de la réalité par l’observateur qui rapporte la donnée et donc la situation. Ce défi est peut-être le plus important dans la chaîne d’analyse car il se situe à la source de la donnée.

Une fois la donnée récoltée il faut être capable de l’interpréter de manière pertinente pour pouvoir prendre la meilleure décision. Certaines informations digitales comme les historiques de transaction, le nombre de crédits en défaut de paiement, le nombre de bons d’achat activés par téléphone sont la définition d’une donnée brute. Cela dit, bien que déterminer si oui ou non une donnée fiable est assez systématique, l’interpréter n’est jamais chose facile.

Un terme qui revient souvent est le biais de sélection à l’échantillonnage. Le fait que des gens utilisant un téléphone qui génère des données pouvant être analysées en temps réel peut poser certains problèmes dans la mesure où cet échantillon d’individus n’est pas représentatif de la population entière. En fonction du type de données, l’observateur peut s’attendre à un groupe plus jeune ou vieux, à une population plus riche ou plus pauvre, plus éduquée ou moins ; plus masculine ou féminine. Par exemple : l’analyse d’un ensemble de recherche Google sera probablement plus représentative de personnes jeunes que de personnes âgées. Ce sont ces attributs qui sont propres aux individus et qui influencent leurs comportements qui font que l’échantillonnage se retrouve biaisé. Les jeux de données issus des méthodes plus traditionnelles comme les sondages sont moins assujettis à ce type de problèmes (un sondage par exemple). Le problème de ces jeux de données est qu’ils ne pourront jamais être généralisés au-delà de l’échantillon. Prenons l’exemple d’un village où une ligne téléphonique recevant des appels reportant des cas de choléra reçoit deux fois plus d’appels que d’habitude juste après une hausse des températures. Parce que la population possédant un téléphone n’est pas représentatif de l’ensemble du village, il est impossible de dire si le nombre de personnes atteintes du choléra a vraiment doublé dans la communauté.

De plus, avec cette quantité massive de données générées il y a un risque de se concentrer uniquement sur la recherche de corrélation et faire des conclusions hâtives sans réelle compréhension de la situation. Cette quantité de données massive peut aussi permettre à certains spécialistes de trouver de nouvelles corrélations alors que ces dernières sont des corrélations évidentes.

On peut penser au tremblement de terre d’Haïti qui a fêté ces dix ans au mois de janvier 2020. Une corrélation a été trouvée entre les dégâts que subissent les bâtiments et le flux de SMS. Cela dit ces derniers étaient corrélés seulement en raison de l’existence d’un bâtiment. En conséquence, la corrélation entre le flux de SMS et les bâtiments endommagés sont artificielles. Il aurait été très risqué pour les organismes de secours de se baser sur ces informations pour déployer leurs unités de sauvetage.

### 6.4 Détecter les anomalies

De manière générale, quand il s’agit de détecter ou mesurer une anomalie dans un écosystème humain est la caractérisation d’une anomalie. Détecter une anomalie socio-économique est très différent que de détecter une anomalie dans le système d’un moteur ou de tout autre système dynamique complexe. Que l’écosystème soit petit ou grand (à l’échelle d’un foyer ou d’un village) il est très difficile de déterminer exactement quelle sont les données d’entrées et de sorties à prendre en compte. Plusieurs exemples montrent l’importance de la spécificité et la sensibilité dans l’étude

**Le Big Data au service d'un développement socio-économique**

d'un système. La sensibilité fait référence à la la capacité d'un système à détecter tout les cas pour lesquels il a été conçu. Quant à la spécificité, elle fait référence à la capacité à détecter seulement les cas recherchés. Le premier cas sont des erreurs dites de Type I communément appelé la proportion de faux positifs alors que le second traite les faux négatifs et fait référence aux erreurs de type II. Une de ces erreurs sera toujours minimisée au détriment de l'autre. Privilégié une erreur plutôt qu'une autre dépendra toujours du système étudié il déterminé laquelle sera la plus utile sera déterminante pour renseigner sur la performance des futures algorithmes.

## Le Big Data au service d'un développement socio-économique

### 7. CONCLUSION

Pour conclure ce travail de recherche, nous avons vu les possibilités proposées par le Big Data sont multiples grâce au développement de ces technologies. Si le développement de ces industries permet de nouvelles possibilités, il faut qu'il y ait une volonté politique et sociale qui soutiennent ces initiatives et que des moyens soient mis en place pour permettre aux pays en voie de développement de tirer profit au maximum des données. Cela dit il ne faut pas reposer l'aide au développement entièrement sur les données, car même si ces dernières ont un fort potentiel, elles rencontrent certaines limites, techniques, politiques mais aussi sociales. Les questions de vie privée resteront toujours centrales et les populations se doivent de s'approprier leurs données pour permettre aux experts et aux spécialistes de l'aide au développement de pouvoir contribuer au mieux pour le bien-être des populations.

Enfin, il ne faut pas tomber dans le piège des technologies qui ont le don de nous faire croire qu'elles peuvent résoudre tous les maux du monde d'aujourd'hui. Il faut que les acteurs du développement aient raison garder et les inégalités entre les différentes parties du monde ne seront pas effacées grâce au Big Data. Ces pays ont des paliers à franchir en termes d'institution, d'éducation, d'infrastructure. Si le Big Data peut servir dans des situations de crises et donner des indicateurs pour prendre des meilleures décisions, il faut une volonté politique du côté des preneurs de décisions dans ces pays pour permettre à ces derniers de se développer de manière pérenne et durable. Il ne faut pas oublier le rôle de la politique étrangère des puissances mondiales qui font passer leurs intérêts géopolitiques et économiques au premier plan, au détriment de certaines populations. C'est pour cela que ce n'est qu'avec une cohésion de partage, de savoirs, et une volonté commune de vouloir contribuer au développement entre pays du Nord, pays du Sud, et organismes internationaux que le Big Data pourra contribuer au développement.

figure 6 - Help from the root



source: *Human hand print tree concept for social* [Tree made of colourful hands s.d]



## Le Big Data au service d'un développement socio-économique

### Référence et Bibliographie

#### Références dans le texte

- 1: <https://upnumerique.com/>
- 2: Site web renseignant sur les statistiques de google <https://www.internetlivestats.com/google-search-statistics/>
- 3: <https://www.collinsdictionary.com/dictionary/english/pattern>
- 4: [https://www.francetvinfo.fr/monde/afrique/en-afrique-le-boom-du-telephone-portable-dope-l-economie\\_3070913.html](https://www.francetvinfo.fr/monde/afrique/en-afrique-le-boom-du-telephone-portable-dope-l-economie_3070913.html)
- 5: <https://www.vodafone.com/what-we-do/services/m-pesa>
- 6: <http://www.rfi.fr/fr/afrique/20170104-reseaux-sociaux-facebook-tunisie-technologie-connectes-continent-africain-twitter>
- 7: Quand le capitalisme perd la tête - **Joseph E. Stiglitz**
- 8: Crise alimentaire de 2011, Corne de l'Afrique [https://fr.wikipedia.org/wiki/Crise\\_alimentaire\\_de\\_2011\\_dans\\_la\\_Corne\\_de\\_l'Afrique](https://fr.wikipedia.org/wiki/Crise_alimentaire_de_2011_dans_la_Corne_de_l'Afrique)
- 9: "Economy: Global Shocks to Become More Frequent, Says OECD." Organisation for Economic Co-operation and Development. 27 June. 2011.
- 10: Fuentes, Nieva Ricardo, and Papa A. Seck. Risks, Shocks and Human Development: On the Brink. Basingstoke, England: Palgrave Macmillan, 2010
- 11: Fonds, programmes et institutions des Nations Unies <https://www.un.org/fr/sections/about-un/funds-programmes-specialized-agencies-and-others/>
- 12: La donnée l'or noir du 21ème siècle <https://www.linkedin.com/pulse/la-donn%C3%A9e-lor-noir-du-21%C3%A8me-si%C3%A8cle-tarek-mansour>
- 13: Définition du crowd-sourcing [https://fr.wikipedia.org/wiki/Production\\_participative](https://fr.wikipedia.org/wiki/Production_participative)
- 14: <https://www.unglobalpulse.org/>
- 15: Le Big Data à découvert (cf. Bibliographie) chapitre 1
- 16: Expérience de pensée tirée de **Big Data for development: Challenges and Opportunities** (cf. Bibliographie)
- 17: Le Big Data à découvert (cf. Bibliographie)
- 18: <https://www.unglobalpulse.org/>
- 19: Big Data for Development : <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobaIPulseMay2012.pdf>
- 20: UN Global Pulse, 'Using Big Data Analytics for Improved Public Transport,' Project Series, no. 25, 2017.

#### Livres:

- Mokrane Bouzeghoub, Rémy Mosseri **Les big data à découvert**, Paris: CNRS Editions, 2017, 364 p.
- Mokrane Bouzeghoub, Rémy Mosseri **Big data, open data: quelles valeurs? Quels enjeux?** Paris : ADBS, DL 2015, 282 p.
- Lafrate Fernando, **Du big Data au smart data : au service d'un monde connecté**, London : ISTE editions, 2015, 86 p.
- Ali Kharrazi Hua Qin and Yi Zhang **Urban Big Data and Sustainable Development Goals: Challenges and Opportunities** Academic Editor: Tan Yigitcanlar ; Published: 9 December 2016 7 p.

#### Travaux de recherches:

- UN Global Pulse (May 2012) **Big Data for Development: Challenges and Opportunities**.
- UN Global Pulse, 'Using Big Data Analytics for Improved Public Transport,' Project Series, no. 25, 2017.

#### Conférences:

- **Symposium on Big Data and Human Development by Professor Bitange Ndemo**, <https://www.youtube.com/watch?v=N7exti9weel>

## **Le Big Data au service d'un développement socio-économique**

Images:

[Tree made of colorful hands s.d]

Disponible sur <https://www.dreamstime.com/human-hand-print-tree-concept-social-help-tree-made-colorful-human-hands-branches-community-help-concept-diverse-culture-image119810671>