

Quel usage du big data dans les élections présidentielles ?

Article paru le 13 avril 2017 sur Manageo

<https://www.manageo.fr/blog/data-bca1/quel-usage-du-big-data-elections-presidentielles-bar13854.html>

Rédigé par Julie Girard

Il ne reste que quelques jours avant le premier tour de l'élection présidentielle française. Plus que quelques jours pour convaincre ou faire changer d'avis un maximum d'électeurs.

Outre les meetings et la présence dans les médias qui se multiplient, les candidats et leurs équipes de campagne recherchent plus que jamais à cibler les bonnes personnes. Et pour y arriver, ils n'hésitent plus à s'appuyer sur le big data.

Zoom sur ces techniques de big data électoral désormais devenues monnaie courante dans l'univers politique.

Le premier objectif en politique est d'arriver à convaincre les électeurs et donc de cibler les bonnes personnes. Le porte-à-porte est très certainement la meilleure manière de convaincre un électeur indécis ou, mieux encore, de le faire changer d'avis.

D'après Alan Gerber et Donald Green, la proximité créée grâce au porte-à-porte permet de « *faire changer d'avis un électeur sur 14, contre un sur 38 au téléphone et un sur 100 000 par email !* ».

On pourrait croire que faire du porte-à-porte est assez simple et ne nécessite pas de préparation particulière : il suffit d'avoir une équipe de militants motivée qui part à la conquête des électeurs en sillonnant des rues ou des quartiers entiers et frappe aux portes de centaines voire de milliers d'électeurs.

Mais depuis plusieurs années, les équipes ne toquent plus aux portes par hasard. Tout est minutieusement préparé et étudié et le porte-à-porte se fait désormais de manière totalement ciblée et calculée.

Liegey Muller Pons (LMP), une startup française spécialisée dans la stratégie électorale, a accompagné François Hollande lors de l'élection présidentielle en 2012, Anne Hidalgo à l'occasion de l'élection municipale de Paris en 2014 ou encore Emmanuel Macron pendant sa Grande Marche de l'été 2016.

Dis-moi où tu habites, je te dirai pour qui tu vas voter

Cette entreprise a développé un logiciel, 50+1, qui permet d'agréger différentes bases de données : des données démographiques issues de l'INSEE ainsi que des données tirées des élections précédentes. Le but est très clair : identifier le vote probable de certaines zones géographiques.

Guillaume Liégey, l'un des cofondateurs de LMP résume d'ailleurs bien cette stratégie de ciblage électoral : « Si vous me dites où vous habitez, je peux vous dire pour qui vous aller voter. ».

Les équipes des candidats savent ainsi vers quels villes, quartiers ou même immeubles il est le plus stratégique de se tourner s'ils veulent convaincre un maximum d'électeurs.



L'équipe de François Hollande, qui a utilisé les services de LMP en 2012, aurait ainsi frappé à près de 5 millions de portes et réussi à convaincre 1 électeur sur 5 qui s'apprêtait à voter pour le Front National de finalement voter pour leur candidat.

Une stratégie de ciblage électoral qui semble donc porter ses fruits.

Cet été, Emmanuel Macron avait également sollicité Liegey Muller Pons à l'occasion de sa Grande Marche.

Le but des militants d'En Marche était de collecter un maximum de verbatim pour comprendre les problèmes concrets des Français. 6 000 volontaires avaient alors été mobilisés. Point de départ de ce sondage géant, l'identification des segments de population et les préoccupations propres à chacun. Il a fallu ensuite cibler les électeurs ou sympathisants potentiels qu'il était pertinent d'interroger.

Une fois ces milliers de verbatim collectés, ils ont été analysés par [Proxem](#), une autre startup française, experte dans le big data et la sémantique.

Ils ont permis au mouvement d'Emmanuel Macron d'identifier les problématiques des Français en fonction des différents profils interrogés et d'adapter le programme et les propositions pour toucher le plus grand nombre de citoyens.

D'autres candidats devraient également utiliser prochainement les services de Liegey Muller Pons puisque pour les prochaines élections législatives, il semblerait que près de 300 candidats soient en passe d'utiliser ce logiciel !

A l'occasion de la primaire de la droite à l'automne 2016, un autre outil numérique de ciblage électoral avait été créé, [l'application Knockin](#). Le but était simple : identifier les sympathisants de droite et faire du porte-à-porte pour les convaincre de voter pour Nicolas Sarkozy.

L'application permettait de cibler les sympathisants de l'ancien Président de la République en croisant toutes les données publiques (listes électorales, réseaux sociaux...) qu'elle avait à disposition. Les équipes n'avaient ensuite plus qu'à cibler ces électeurs en priorité.

La France n'est pas un pays précurseur dans le ciblage électoral. Outre-Atlantique, on peut citer les campagnes de Barack Obama ou, plus récemment, celle de Donald Trump lors de la dernière campagne présidentielle.

D'après CNN, le nouveau Président des Etats-Unis aurait investi près de 5 millions de dollars dans la Data Management Platform (DMP) Cambridge Analytica. Cet outil était habituellement plutôt utilisé par les publicitaires dans le domaine du marketing pour mesurer les performances de leurs campagnes.

La mécanique était assez simple : tout commençait avec un test de personnalité sur des réseaux sociaux comme Facebook. Les données issues de ces tests permettaient de mieux cibler les électeurs potentiels et de comprendre leur profilage psychologique. Combiné avec d'autres données gratuites ou payantes, comme les lieux habituels de shopping ou encore les habitudes électorales de ces citoyens, l'algorithme de la DMP permettait alors de proposer à ces internautes des publicités pro-Trump ciblées en fonction de leur profil.

Certaines équipes, à l'instar de celles de Jean-Luc Mélenchon ou François Fillon, ne s'arrêtent pas là et vont encore un peu plus loin dans l'usage des outils numériques dans le cadre de leurs campagnes.

Les deux ont fait appel à [NationBuilder](#), une plateforme américaine qui permet à tout type d'entités (entreprises, partis politiques, ONG...) de mener des campagnes électorales.

Les références de cette entreprise sont internationales et issues d'univers très différents : Airbnb, les Nations Unies, Amnesty International, l'équipe de Donald Trump, des partis politiques du Royaume-Uni ou encore australiens...

Toni Cowan-Brown, vice-présidente de la société, a expliqué à nos confrères [d'Usbek & Rica](#) que la promesse de NationBuilder était « *d'aider [ses] clients à implémenter les meilleurs outils qui existent, pour que leur campagne soit la meilleure possible. [...] L'idée, c'est vraiment de pouvoir offrir à tout le monde les outils nécessaires pour lancer une campagne, et potentiellement la gagner. Qu'on soit François Fillon, Jean-Luc Mélenchon ou un candidat dont personne n'a encore entendu parler, on peut avoir accès aux mêmes outils. C'est quelque chose que je trouve fascinant, parce que ça démocratise les outils politiques.* ».

Cette plateforme permet aux partis de créer des bases de données riches et précises, de constituer des fichiers de militants pour les relancer fréquemment ou encore d'identifier les zones géographiques sur lesquelles il faudrait plus investir.

NationBuilder propose donc un site puissant qu'ils gèrent entièrement : la firme développe le site, gère les bases de données, envoie les emails, tout cela sur une seule et même plateforme.

Toni Cowan-Brown précise l'intérêt d'un tel site « tout-en-un » : « *Quelqu'un peut venir sur le site, entrer son code postal et dire qu'il veut se joindre à un groupe local. L'information est alors renvoyée à la base de données, avant d'être adressée à l'équipe de campagne, qui peut contacter cette personne et lui proposer d'agir.* ».

C'est ainsi que le mouvement de la France insoumise a créé 3 200 groupes d'appui à travers toute la France, chacun réunissant entre 3 et 15 personnes, pour aller appuyer la candidature de Jean-Luc Mélenchon.

Les équipes précisent cependant que les outils ayant énormément augmenté pendant la campagne, « *NationBuilder ne gère aujourd'hui plus que la partie groupes d'appui et évènements locaux.* ».

Les équipes de campagne ne s'arrêtent pas au ciblage électoral. Il est primordial pour elles de savoir si toutes les actions qu'elles mènent influent sur les intentions de vote et elles étudient donc attentivement les sondages. Dans ce domaine, les sondages déclaratifs ne sont plus seuls, le datamining prend désormais une part de plus en plus importante.

L'importance du buzz

Jusqu'à récemment, les sondages étaient réalisés à partir d'échantillons de population, sur la base de données déclaratives. Aujourd'hui, ils viennent s'enrichir de nombreuses données digitales issues des réseaux sociaux.

L'idée est de mesurer le poids de chaque candidat sur les réseaux sociaux en récupérant un maximum de posts, tweets et autres messages et de se servir d'algorithmes d'analyse lexicale pour les analyser.

Il est ainsi très facile de mesurer le buzz (grâce à la quantité de messages postés) et l'émotion (via la qualité des posts) associés à chaque candidat et de déterminer l'impact que cela pourrait avoir au moment du vote pour affiner les prévisions.

L'important ici n'est pas de savoir comment nous parlons d'un candidat et si celui-ci est perçu avec une opinion plutôt positive ou négative. Ce qui compte, c'est simplement que l'on parle de lui.

Cette analyse des données digitales présente plusieurs avantages par rapport aux sondages traditionnels.

La base étudiée est déjà beaucoup plus large qu'un panel traditionnel. Et comme il n'y a pas de question posée, il n'y a ni orientation dans la réponse ni hésitation à répondre ce que l'on pense « réellement ».

Le datamining permet donc d'être au plus près des émotions ressenties et de mieux capter les tendances en temps réel.

[Filteris](#), une société canadienne spécialisée dans l'analyse de données sur le web, se positionne déjà depuis plusieurs années en solution complémentaire aux sondages plus traditionnels.

En 2012, à l'occasion du premier tour de l'élection présidentielle française, elle anticipait déjà le podium du 22 avril : François Hollande en tête, Nicolas Sarkozy en deuxième position suivi de Marine Le Pen.

Plus récemment, en octobre 2016, elle avait anticipé l'ascension de Donald Trump.

Avec différents propos qu'il avait tenus, il était au cœur de nombreuses polémiques. Et même s'il ne recueillait que 21% d'avis positifs, ce qui comptait, c'était qu'on parle de lui et le buzz qu'il créait. Filteris lui avait alors donné un poids numérique plus important que celui attribué à Hillary Clinton, avec le succès que l'on connaît par la suite.

La CNIL, le gendarme de la data

Mais la récolte de toutes ces données n'est pas sans limite. La Commission nationale informatique et libertés, la CNIL, surveille de près ces nouveaux outils numériques.

A l'occasion de la primaire de la droite et du centre, elle avait d'ailleurs tenu à rappeler que « le « crawling » des réseaux sociaux par les logiciels, aux fins de collecte et de traitement de données disponibles publiquement, n'est pas légal en l'absence d'information des personnes. » et que « les personnes qui ont volontairement fourni leur adresse électronique aux fins de recevoir une newsletter de tel candidat ne peuvent être considérées comme ayant été informées ou ayant consenti à nouer des relations avec ce candidat par le biais d'un réseau social. ».

NationBuilder a ainsi désactivé une de ses fonctionnalités qui consistait, à partir d'un simple email, à enrichir ses bases de données avec des informations issues des réseaux sociaux.