

Big data et marketing politique

La controverse étudiée s'intéresse à l'usage des sciences des données à des fins politiques et à l'émergence de nouvelles méthodes de marketing politique. Plusieurs questions sont soulevées par cette controverse et nous nous intéresserons à certaines d'entre elles et aux débats qu'elles posent dans l'arène scientifique et technique.

1 Comment les campagnes électorales obtiennent-elles des données ?

De plus en plus de campagnes électorales modernes collectent des données sur les citoyens éligibles de voter pour leur candidat.

Une des plus importantes ressources auxquelles les campagne ont accès sont les données *publiques* collectées par des institutions gouvernementales [13]. De là, les campagnes obtiennent des informations comme le nom, l'adresse, le numéro de téléphone et la participation aux précédentes élections d'électeurs inscrits sur les listes électorales. Elles ajoutent à cela des données récoltées par des recensements récents pour avoir plus d'informations sur le niveau de revenu, d'éducation et l'origine ethnique(US) des électeurs. Et elles obtiennent la liste publique(US) des citoyens ayant contribué financièrement à des campagnes passées. Ces données constituent la majorité des données dont disposent les campagnes électorales. Elles génèrent également leurs propres données en faisant des sondages, des porte-à-porte et en collectant des données depuis leur site web de campagne.

Depuis quelques années, les campagnes électorales ne se limitent plus aux bases de données publiques. Elles achètent des données de consommateurs [7] fournies par des *data brokers* comme Axciom ou Epsilon qui collectent énormément de données comme les achats, les abonnements, le comportement en ligne (grâce au cookie tracking) et les habitudes de visionnage de TV sur des millions de consommateurs.

Débat : Certains journalistes et avocats dénoncent ces pratiques qu'ils considèrent être une atteinte à la vie privée des citoyens. Le caractère publique de la majorité des données collectées sert de défense aux campagnes électorales [13]. Et pour ce qui est des données privées, Epsilon dispose d'un brevet [6] pour une méthode qui permettrait de faire une correspondance *double-blind* entre des bases de données tout en respectant les lois régissant les données privées.

Ces données sont stockées dans de larges bases de données baptisées *Voter Vault*(*Republicains 1990s*), *Demzilla*(*Democrats 2002-2004*), *Narwhal*(*Obama 2012*), *Alamo*(*Trump 2016*). La gestion de celles-ci est facilitée par des entreprises qui fournissent des logiciels de gestion de campagnes comme **NationBuilder** et **NGP VAN** et des boîtes de conseil spécialisées comme **Blue Labs**, **Aristotle** ou **Cambridge Analytica**.

Debat : Ces entreprises privées vantent la quantité de données qu’elles peuvent fournir sur chaque électeurs [2, 3]. Pour elles, il faut chercher à avoir le plus de données possibles. Il existe plusieurs articles qui questionnent l’utilité de cette accumulation de données. Les chercheurs en science politique notent la difficulté de mesurer la véritable valeur ajoutée de ces données et les statisticiens insistent sur la nécessité de construire des modèles sur des données ”de bonne qualité” [12] et non pas n’importe lesquelles pour ne pas être induit en erreur [1] .

2 Comment les données sont-elles utilisées ?

Les données obtenues sont utilisées pour construire des modèles statistiques. Ces modèles sont utilisés pour faire du *micro-targeting*.

2.1 A/B Testing

Les campagnes peuvent affiner le langage, les formulations utilisées dans leurs messages, les images, et même la couleur des boutons de leur site en menant des expériences de terrain nommé A/B testing et en mesurant le taux de succès des différentes variations testées. Elles identifient ainsi quels sujets, quelles formulations et quels arguments parlent le plus à un certain profil d’électeur pour mieux le viser [5].

2.2 Modèle uplift [Campagne Obama 2012]

Pour augmenter l’efficacité et l’allocation des ressources, la campagne d’Obama s’est servi [9] d’un modèle statistique [15] pour identifier les électeurs qui voteront pour Obama **si et seulement si** un contact est réalisé (porte-a-porte, courrier, pub TV, email,,).

2.3 Psychometrics/Psychographic modeling [Cambridge Analytica]

Il existe des articles détaillant des méthodes de classification de la personnalité des individus à partir de likes facebook et d’autres données [11, 14]. Cette information supplémentaire sur les électeurs a été utilisée pour cibler des types de personnalité par Trump en 2016 [17]. Cette méthode a été enrichie par les méthodes d’élargissement d’audience de Facebook qui détermine des profils similaires à l’ensemble de profils initialement visés.

Debats : Ces méthodes et modèles mathématiques qui incluent également des méthodes de régressions linéaires non mentionnées ne sont pas véritablement contestés. Elles sont largement utilisées dans le monde du marketing. Les questions se trouvent plutôt autour de leur place et de leur efficacité dans le monde politique.

3 Quelle est la réalité de leur efficacité sur les électeurs ?

L'argent dépensé dans les efforts de micro-targeting témoigne de la foi qu'ont les campagnes électorales en ces méthodes. Dans le monde académique, certains chercheurs montrent qu'un contact personnalisé avec les électeurs est bénéfique pour encourager le vote [18]. D'autres remarquent que le micro-targeting peut être interprété comme un manque d'authenticité [8] et que le ciblage peut détourner des électeurs [16] surtout si des erreurs de ciblage sont commises. La difficulté de mesurer l'efficacité du ciblage [10] fait que le monde académique n'est pas tranché sur la question du ciblage en politique, contrairement au monde politique qui investit de plus en plus dans le big data [4].

References

- [1] 15 Insane Things That Correlate With Each Other.
- [2] Aristotle Political Technology to Power Democracy.
- [3] HaystaqDNA.
- [4] This Is How Facebook Actually Won Trump the Presidency.
- [5] How Obama Raised \$60 Million by Running a Simple Experiment, November 2010.
- [6] Lois Beckett. How Microsoft and Yahoo Are Selling Politicians Access to You, June 2012.
- [7] Author: Kim Zetter Kim Zetter Business. For Sale: The American Voter.
- [8] Eitan D. Hersh and Brian F. Schaffner. Targeted Campaign Appeals and the Value of Ambiguity. *The Journal of Politics*, 75(2):520–534, April 2013.
- [9] Sasha Issenberg. How Obama’s Team Used Big Data to Rally Voters.
- [10] David Karpf. The Internet and American Political Campaigns. *Forum-a Journal of Applied Research in Contemporary Politics*, 11(3):413–428, October 2013. WOS:000325925200007.
- [11] Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, April 2013.
- [12] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6176):1203–1205, March 2014.
- [13] David W. Nickerson and Todd Rogers. Political campaigns and big data. *The Journal of Economic Perspectives*, 28(2):51–73, 2014.
- [14] Julie Hassing Nielsen. Personality and Euroscepticism: The Impact of Personality on Attitudes Towards the EU. *Jcms-Journal of Common Market Studies*, 54(5):1175–1198, September 2016. WOS:000388838500011.
- [15] Piotr Rzepakowski and Szymon Jaroszewicz. Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*, 32(2):303–327, August 2012.
- [16] Ali A. Valenzuela and Melissa R. Michelson. Turnout, Status, and Identity: Mobilizing Latinos to Vote with Group Appeals. *American Political Science Review*, 110(4):615–630, November 2016. WOS:000392739900001.
- [17] Joel Winston. How the Trump Campaign Built an Identity Database and Used Facebook Ads to Win the Election, November 2016.
- [18] “Don’t Talk to Strangers: Experimental Evidence of the Need for Targeting.” Presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 12- 15, 2007.