

Datagotchi : Les élections générales québécoises de 2022

Annexe – Description du projet

2022-04-21

La création de la plateforme web Datagotchi pour les élections générales canadiennes de 2021¹ et la collecte de données dans le cadre de ce projet ayant été réalisée avec succès, la Chaire de leadership en enseignement des sciences sociales numériques (CLESSN) souhaite proposer une nouvelle édition améliorée de cette application. Cette fois-ci, les électeurs pourront explorer la façon dont leurs caractéristiques reliées à leur mode de vie permettent à l'équipe de recherche de faire une prédiction quant à leur choix de vote aux élections générales québécoises de 2022.

État des connaissances et théorie

Avancées technologiques obligent, il est aujourd'hui plus facile que jamais d'accumuler et de traiter des informations de toute sorte. Il n'y a virtuellement plus de limite à la quantité d'information que nous pouvons traiter à l'aide d'ordinateurs. Nous sommes bel et bien entrés dans l'ère numérique. Bien que les quantités massives de nouvelles données accessibles représentent une mine d'or informationnelle pour les chercheurs en sciences sociales, la numérisation quotidienne de la vie privée soulève toutefois son lot de critiques légitimes. Certains événements récents illustrent les débats éthiques soulevés par de telles pratiques — le scandale lié à Facebook et à la firme Cambridge Analytica en est un exemple phare. Néanmoins, tout porte à croire que ces pratiques perdureront, et nous pouvons donc nous interroger sur leur contribution. Quelles sont les contributions potentielles de ces nouvelles données numériques à notre compréhension des phénomènes sociaux ?

Ce projet s'attaque à une problématique théorique précise. Trop de projets liés aux données numériques massives sont abstraits et ne s'attardent qu'aux aspects techniques de la question. L'objectif de ce projet est de mettre les données au service de la connaissance scientifique humaine. Une théorisation claire doit guider la recherche. À cet effet, tandis que le pouvoir prédictif et explicatif des variables sociodémographiques traditionnelles — telles que la religion, le revenu et l'éducation — est en baisse, il apparaît essentiel de repenser la manière dont les études d'opinion publique sont conduites. Les données numériques massives constituent un pilier de cette réorientation des méthodologies de recherche. Elles permettent désormais d'étudier différents phénomènes avec une précision temporelle et géographique granulaire (Lazer et Radford, 2017). Dans le même ordre d'idées, State, Park, Weber et Macy (2015) et de Beauchamp

¹ Datagotchi a reçu l'approbation du Comité éthique de l'Université Laval le ... 2021 (numéro d'approbation : 2021-061/21-07-2021).

(2017) ont examiné des réseaux d'interaction impliquant des millions de personnes sur Twitter et sur Yahoo afin de faire des prédictions électorales autrefois impensables. Les données recueillies ont notamment permis de révéler des « corrélations intra- culturelles » profondes entre des individus distribués un peu partout sur la planète. Ces recherches prometteuses illustrent dans quelle mesure les données massives numériques permettent la construction d'une nouvelle « science de la société » ayant pour objectif d'étudier la société comme entité entière composée de sous-systèmes et d'individus connectés de façon dynamique selon leurs « manières d'être » et leur location géospatiale (Lazer et Radford, 2017 : 25).

Ces recherches permettent également de constater un morcellement des pratiques traditionnelles additionné à une diversification croissante des intérêts et des préférences individuelles (Lamont, Schmalz-Bauer, Waller et Weber, 1996 : Fischer et Mattson, 2009). En ce sens, la réduction importante de la capacité explicative et prédictive des variables sociodémographiques conventionnelles pose un problème à la fois inquiétant et stimulant. Alors que les études quantitatives contemporaines en science politique font généralement appel à des données de sondages publics et électoraux, il apparaît aujourd'hui nécessaire d'élargir et d'approfondir ces bases de données en y intégrant des marqueurs de socialisation diversifiés. La fragmentation croissante de la sphère culturelle et sociale — illustrée notamment par l'hyperspécialisation des intérêts, des préférences, des goûts et des connaissances — peut contribuer à expliquer cette transformation sociale profonde (Bishop, 2009 : Bertrand et Kamenica, 2018). Tout porte cependant à croire que les variables personnalisées concernant le style de vie pourraient permettre aux modèles de prédiction électorale de regagner leur pouvoir explicatif et prédictif. Les possibilités d'accumulation d'un volume important de données permettent par ailleurs de concevoir des analyses sophistiquées avec un pouvoir de précision impressionnant. Ce projet de recherche s'insère au cœur de ces possibilités nouvelles. Effectivement, alors que les premiers travaux empiriques en opinion publique questionnaient l'existence même des attitudes politiques individuelles (Converse, 1964: Zaller, 1992), des recherches plus récentes démontrent que des outils de mesure plus raffinés arrivent à cerner les attitudes latentes (Stimson, 2018).

Certains outils permettant de mesurer les attitudes sur des enjeux éthiques et moraux ont, par exemple, été développés et validés (Ansolabehere, Rodden et Snyder, 2008). L'étude des valeurs et des croyances profondes joue ainsi un rôle crucial dans la recherche contemporaine en opinion publique et en psychologie politique (Feldman, 1988: Inglehart, 1990 : Jacoby, 2006). Enfin, la diminution marquée de la capacité prédictive des modèles traditionnels combinée aux possibilités exponentielles issues des nouvelles données numériques ont mené certains chercheurs à s'intéresser aux « *lifestyle politics* » (Giddens, 1991 : Bennett, 1998: Micheletti, 2003). En effet, la fragmentation des unités de socialisation — style de musique, allégeance esthétique ou artistique, mode de vie quotidien, style vestimentaire, habitudes alimentaires, etc. — pourrait aider à expliquer les valeurs, les préférences, les attitudes et les opinions individuelles (Weiss, 1988 : Chaney, 2002: Stolle, Hooghe et Micheletti, 2005). Les travaux de DellaPosta, Shi et Macy

(2015) et de Mutz et Rao (2018) sur les raisons expliquant l'attachement des progressistes américains aux *cafés lattés* en sont un exemple. Ces variables sociales, en perpétuelle mutation, demeurent toutefois peu explorées dans les études en opinion publique et en psychologie. Leur intégration promet des avancées scientifiques théoriques et empiriques prometteuses.

Objectifs

Les objectifs centraux du projet demeurent les mêmes que dans la première édition, bien que l'interface visuelle et l'expérience utilisateur seront améliorées.

(1) Il y a d'abord un objectif scientifique. Datagotchi permet aux chercheurs d'étudier la relation entre les habitudes de vie et les attitudes et comportements politiques. Un des postulats de base est que les habitudes de vie sont de plus en plus d'intérêt afin de comprendre différents comportements sociaux et politiques, dont le choix de vote. Certaines caractéristiques liées aux habitudes de vie (par exemple, les habitudes de consommation, les loisirs, les préférences musicales et de film) sont de plus en plus utilisées par les stratèges des partis politiques, mais demeurent peu intégrées en recherche. Datagotchi est donc un outil de recherche scientifique qui collecte des données de façon confidentielle et sécuritaire, afin de mieux comprendre la relation entre habitudes de vie et comportement électoral.

(2) Via une plateforme accessible, ludique et éducative, démocratiser les outils de ciblage utilisés par les partis politiques et conscientiser les citoyens aux enjeux démocratiques et éthiques soulevés par l'ère numérique et l'utilisation de l'IA par les acteurs politiques. En effet, dans un contexte où les victoires électorales s'obtiennent parfois par un mince pourcentage de voix et où le marketing politique occupe une place croissante, il est important que les citoyens soient informés de la façon dont les données numériques sont utilisées par les partis politiques. Ce projet de recherche, en plus d'utiliser les technologies issues de l'IA afin de mener ses analyses, cherche à sensibiliser les Québécoises et les Québécois à l'utilisation de l'IA à des fins électorales, ainsi qu'aux enjeux éthiques qui en découlent. En ce sens, Datagotchi se veut en quelque sorte un outil de conscientisation et d'éveil face à l'utilisation à grande échelle des données personnelles par les partis politiques.

(3) Enfin, Il y a un objectif clair de vulgarisation scientifique et de transfert des connaissances. L'idée d'une plateforme interactive est aussi de démocratiser l'analyse statistique afin que les utilisateurs puissent explorer, de façon intuitive, l'effet de différentes variables sur les prédictions électorales. L'objectif est par ailleurs de construire une plateforme raffinée sur le plan visuel (et dont l'expérience utilisateur sera améliorée par rapport à la première édition). Grâce à sa facilité d'utilisation, mais aussi grâce à la rigueur scientifique des analyses, la plateforme captera l'intérêt de publics

divers : elle sera attrayante pour des citoyens d'âge scolaire comme pour des chercheurs, des stratèges ou des décideurs publics.

Méthodologie

Infrastructure

La nouvelle édition de Datagotchi sera construite à partir de la même infrastructure que la précédente, même si le visuel sera amélioré (ajout d'animations, raffinement des images, amélioration de l'application pour la version 'Desktop') et que quelques changements seront apportés, afin de simplifier son utilisation (par exemple, modifications du tableau de bord afin de rendre la manipulation des données plus intuitive pour l'utilisateur, etc.).

Activités de collecte

Pour la sélection des questions ajoutées sur le sondage de la plateforme, il y aura deux nouvelles études pilotes en collaboration avec la firme de recherche et marketing Synopsis.

Analyse des données

Différents modèles statistiques seront comparés (par exemple, régression logistique, modèles de forêts d'arbres décisionnels, etc.) afin de celui permettant la prédiction de vote la plus robuste et performante. Ces analyses préliminaires seront réalisées à partir de données collectées par sondage (pilote 1 et pilote 2). À l'aide de ces données préliminaires, un algorithme d'apprentissage sera entraîné, et qui sera en suite en mesure de faire des prédictions sur de nouvelles données (soit celles des utilisateurs de Datagotchi, une fois la plateforme déployée). Au fur et à mesure que les utilisateurs utiliseront Datagotchi, ils pourront nous dire si la prédiction faite par l'application est correcte ou non, ce qui contribuera aussi à améliorer encore l'algorithme de prédiction pour les autres utilisateurs.

Conservation des données

Les données seront conservées dans la banque déjà existante de Datagotchi, qui est sous la responsabilité du professeur Yannick Dufresne, et destinée aux participants ayant consenti à cette banque. De ce fait, la gestion et la conservation des données des participants qui ont donné leur consentement se fera tel que décrit dans le cadre de gestion de la banque de données de Datagotchi (N° d'approbation : 2021-065 CG).

Retombées

L'importance du développement de la recherche théorique sur les fondements et la structure de l'opinion publique et des comportements politiques est indéniable. Une telle recherche aura ainsi des échos dans la sphère de la recherche académique. Les chercheurs en science politique et en psychologie et, plus largement, en sciences sociales y verront certainement une contribution essentielle au développement du champ. Cette

recherche a également un impact public. En cherchant à augmenter la capacité et la précision des modèles quantitatifs de prédiction en science politique, les retombées de ce projet seront importantes pour les décideurs publics, les acteurs politiques, les médias et le grand public.

Plus largement, par la diffusion de la plateforme et la vulgarisation des résultats de recherche, ce projet permettra aux citoyens d'avoir une compréhension plus claire et raffinée des structures culturelles et institutionnelles au sein desquelles ils évoluent quotidiennement. À cet égard, une des visées est de démocratiser, à l'aide d'analyses vulgarisées et rigoureuses, les outils de ciblage utilisés par les partis politiques et de conscientiser les citoyens aux enjeux soulevés par l'ère numérique.

Bibliographie

- Ansolabehere, S., Rodden, J. et Snyder, J. M. (2008). The Strength of Issues : Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting. *American Political Science Review*, 102(2), 215–232.
- Barberá, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J. et González-Bailón, S. (2015). The Critical Periphery in the Growth of Social Protests. *PloS One*, 10(11), 1-15.
- Beauchamp, N. (2017). Predicting and Interpolating State-Level Polls Using Twitter Textual Data. *American Journal of Political Science*, 61(2), 490–503.
- Bennett, W. L. (1998). The Uncivic Culture : Communication, Identity, and the Rise of Lifestyle Politics. *PS : Political Science & Politics*, 31(4), 741–761.
- Bertrand, M. et Kamenica, E. (2018). *Coming Apart ? Cultural Distances in the United States Over Time* (Rapport technique). National Bureau of Economic Research.
- Bishop, B. (2009). *The Big Sort : Why the Clustering of Like-Minded America is Tearing Us Apart*. Houghton Mifflin Harcourt.
- Campbell, Angus. 1960. "Converse, Pe Miller, We and Stokes, de (1960). the American Voter." *Ann Arbor*.
- Chaney, D. C. (2002). From Ways of Life to Lifestyle : Rethinking Culture as Ideology and Sensibility. In *Culture in the Communication Age* (p. 85-98). Routledge.
- Converse, P. E. (1964). The Nature of Belief Systems in Mass Publics. In D. E. Apter (Ed.), *Ideology and Discontent* (pp. 206–261). Free Press of Glencoe.
- Dalton, Russell J. 2007. "Partisan Mobilization, Cognitive Mobilization and the Changing American Electorate." *Electoral Studies* 26 (2). Elsevier: 274–86.
- Dalton, Russell J, and others. 1988. *Citizen Politics in Western Democracies: Public Opinion and Political Parties in the United States, Great Britain, West Germany, and France*. Chatham House Chatham, NJ.
- DellaPosta, D., Shi, Y. et Macy, M. (2015). Why do Liberals Drink Lattes? *American Journal of Sociology*, 120(5), 1473–1511.
- Feldman, S. (1988). Structure and Consistency in Public Opinion : The Role of Core Beliefs and Values. *American Journal of Political Science*, 416–440.

- Fischer, C. S. et Mattson, G. (2009). Is America Fragmenting ? *Annual Review of Sociology*, 35, 435-455.
- Giddens, A. (1991). *Modernity and Self-Identity : Self and Society in the Late Modern Age*. Stanford University Press.
- Inglehart, R. (1990). *Culture Shift in Advanced Industrial Society*. Princeton University Press.
- Jacoby, W. G. (2006). Value Choices and American Public Opinion. *American Journal of Political Science*, 50(3), 706–723.
- Krosnick, Jon A. 1990. "Government Policy and Citizen Passion: A Study of Issue Publics in Contemporary America." *Political Behavior* 12 (1). Springer: 59–92.
- Lamont, M., Schmalzbauer, J., Waller, M. et Weber, D. (1996). Cultural and Moral Boundaries in the United States : Structural Position, Geographic Location, and Lifestyle Explanations. *Poetics*, 24(1), 31-56.
- Lazarsfeld, Paul Felix, Bernard Berelson, and Hazel Gaudet. 1944. "The People's Choice." Duell, Sloan & Pearce.
- Lazer, David, and Jason Radford. 2017. "Data Ex Machina: Introduction to Big Data." *Annual Review of Sociology* 43. Annual Reviews: 19–39.
- Micheletti, M. (2003). Shopping with and for Virtues. In *Political Virtue and Shopping* (pp. 149–168). Springer.
- Mutz, D. C. et Rao, J. S. (2018). The Real Reason Liberals Drink Lattes. *PS : Political Science & Politics*, 1–6.
- State, B., Park, P., Weber, I. et Macy, M. (2015). The Mesh of Civilizations in the Global Network of Digital Communication. *PloS One*, 10(5), e0122543.
- Stimson, J. (2018). *Public Opinion in America : Moods, Cycles, and Swings*. Routledge.
- Stolle, D., Hooghe, M. et Micheletti, M. (2005). Politics in the Supermarket : Political Consumerism as a Form of Political Participation. *International Political Science Review*, 26(3), 245-269.
- Wang, W., Rothschild, D., Goel, S., et Gelman, A. (2015). Forecasting Elections with Non-Representative Polls. *International Journal of Forecasting*, 31(3), 980–991.
- Weiss, M. J. (1988). *The Clustering of America*. Harper Collins.

Wood, Wendy, Carl A Kallgren, and Rebecca Mueller Preisler. 1985. "Access to Attitude-Relevant Information in Memory as a Determinant of Persuasion: The Role of Message Attributes." *Journal of Experimental Social Psychology* 21 (1). Elsevier: 73–85.

Zaller, J. et Feldman, S. (1992). A Simple Theory of the Survey Response : Answering Questions versus Revealing Preferences. *American Journal of Political Science*, 36(3), 579- 616. Zaller, J. R. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University Press.