

Ces votations si prévisibles



Docteur en physique quantique, Sébastien Perseguers s'est mis au machine learning il y a trois ans. © Alain Wicht

Publié le 21.02.2022

Sébastien Perseguers a créé un algorithme qui prédit le résultat des objets soumis au peuple suisse

TAMARA BONGARD

Sciences » Le résultat des dernières votations fédérales n'a pas surpris Sébastien Perseguers. Si ce n'est pour constater à quel point ses prédictions étaient proches de la réalité. Le Fribourgeois a estimé que les Suisses rejetteraient l'interdiction de l'expérimentation animale à 82% (ils l'ont fait à 79,1%), qu'ils diraient non à la suppression du droit de timbre d'émission à 62% (c'est 62,7%), qu'ils refuseraient les mesures en faveur des médias à 57% (le niet a atteint 54,6%) et qu'ils accepteraient l'initiative s'attaquant à la publicité pour le tabac à 59% (le oui a finalement obtenu 56,6%). Ce n'est ni dans une boule de cristal ni dans le marc de café qu'il a vu ces chiffres, mais dans un algorithme qu'il a créé.

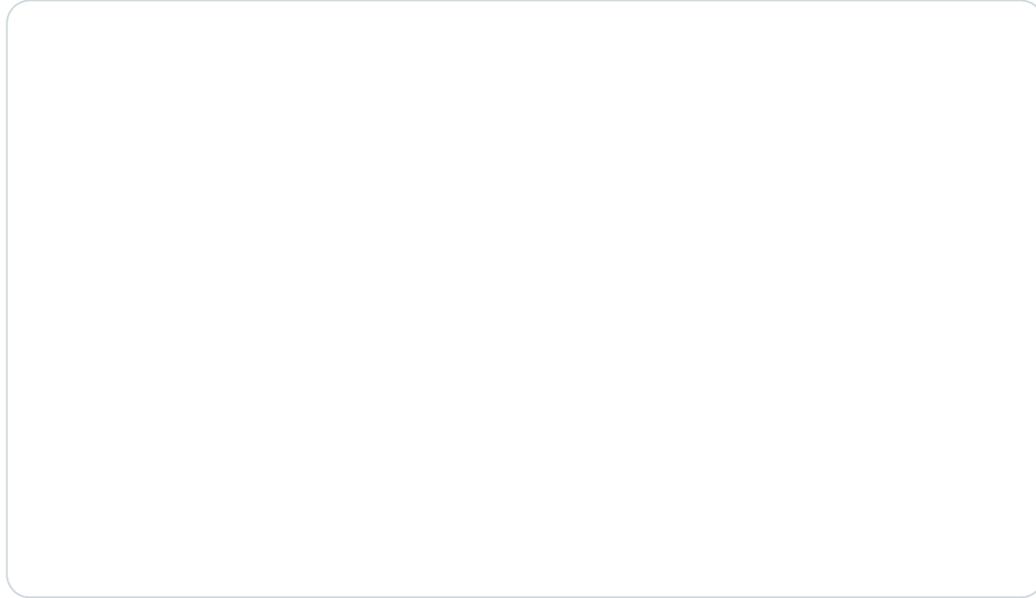


Sébastien Perseguers

@spereguers



Après les résultats officiels #CHvote #13Février, place à la vérification des prédictions #MachineLearning : les 4 résultats ont été parfaitement pronostiqués plus de 10 jours à l'avance, avec une erreur moyenne de seulement 2% !



7:20 PM · 15 févr. 2022



6 Répondre Partager

[Découvrez ce qui se passe sur Twitter.](#)

Ce docteur en physique quantique met gratuitement à disposition ses prédictions sur son compte Twitter, 10 jours avant la sanction des urnes. C'est une manière de montrer un des aspects de son travail, qui est pour le reste marqué du sceau de la confidentialité. Il déploie ses activités dans la société Gradiom, qu'il a fondée il y a six ans. Installée sur le site de BlueFactory, à Fribourg, elle offre un support mathématique aux ingénieurs de grandes entreprises, actives notamment dans l'horlogerie ou dans les systèmes de défense militaire. Il s'est mis il y a trois ans au *machine learning* («l'apprentissage automatique» en français), un outil statistique et numérique pour traiter des données et en extraire une information. En très résumé, il a imaginé sur son ordinateur un modèle qu'il nourrit de chiffres et qui lui donne une réponse à une question, en l'occurrence politique.

Pourquoi vous êtes-vous intéressé aux votations fédérales?

Sébastien Perseguers: Quand j'ai voulu me mettre au *machine learning*, je n'avais pas de projet concret, mais je suis très vite arrivé à la conclusion que les votations pouvaient être un thème intéressant à développer, car des données de qualité étaient accessibles – l'Office fédéral de la statistique est assez merveilleux pour cela. De plus, tous les trois mois elles se répètent, ce qui me permet d'ajuster le modèle au fil du temps. Le sujet est également pertinent car ces prévisions parlent à beaucoup de personnes. Et il se trouve qu'un modèle relativement simple au début marchait étonnamment bien, donc j'ai continué.

Votre marge d'erreur a-t-elle diminué?

Initialement, elle était à plus ou moins 8% alors que les prévisions des deux dernières votations affichent moins de 3% d'erreur en moyenne. Vu la diversité des sujets soumis au vote, le modèle actuel est donc très satisfaisant. D'une part, cela s'explique par les algorithmes que j'utilise, et d'autre part, par l'ajout de données de base.

C'est-à-dire?

Au fur et à mesure, j'ai trouvé de nouvelles sources de données qui apportent de l'information et aident à prédire les résultats avec plus d'exactitude. Le modèle de base pondérait les recommandations de vote de chaque parti politique et leur poids dans chaque canton, ce qui donnait une bonne indication de l'acceptation ou non de l'objet par un canton. En ajoutant encore le résultat des votes au Conseil national et au Conseil des Etats ainsi que les recommandations du Conseil fédéral, j'obtenais une orientation grossière du résultat. Il fallait aussi tenir compte des affinités entre les régions linguistiques et les thèmes abordés, des ajustements selon le type d'objet et de la date de la votation, modifiant parfois légèrement le taux de participation.

Ce sont donc ces informations qui ont permis de réduire la marge d'erreur de l'algorithme?

Pas seulement. L'algorithme signifie que chaque votation est décrite par un certain nombre de paramètres. J'ai cherché comment mélanger au mieux ces données pour expliquer le résultat connu d'une votation particulière. Je me suis basé sur un historique de dix ans, ce qui représente 100 objets de vote.

Ensuite, il y a un caractère plus subtil à décrire mathématiquement. C'est là où j'ai trouvé d'autres sources, tout d'abord les sondages Tamedia et SSR qui permettent d'intégrer un ressenti plus subjectif de la population. J'ai aussi ajouté les analyses faites par l'Université de Zurich dans l'*Abstimmungsmonitor*, qui observe si les médias utilisent un vocabulaire positif ou négatif pour évoquer une votation et qui détermine la tonalité générale des articles. Ma dernière source vient de cantons comme Genève et Neuchâtel indiquant le taux de participation des citoyens une vingtaine de jours avant les votations. Cela m'aide aussi à prédire la participation finale plus précisément, une variable qui peut modifier légèrement le rapport de forces des votants. Ce mélange de paramètres objectifs et subjectifs nuance ma prédiction.

Vos sources «subjectives» suffisent-elles à intégrer les changements de société?

Certains changements se voient déjà au niveau de la pondération par parti. Je me base sur les élections qui ont lieu tous les quatre ans. Ces dernières huit années, la représentation des Verts et des Vert'libéraux a augmenté, et par conséquent la pondération de leur recommandation a été automatiquement incluse dans le modèle. Plusieurs facteurs interviennent, mais il n'y a pas eu d'autres modifications importantes, c'est assez stable... comme la Suisse.

«Mon modèle est largement perfectible»
Sébastien Perseguers



Est-il possible d'améliorer votre modèle?

Il est largement perfectible, mais pour ça il me faudrait du temps et des financements. Dans le *machine learning*, la collecte des données prend un temps phénoménal. A moins d'avoir les moyens humains et financiers des géants du numérique, chaque étape doit être optimisée. Ainsi, moins il y a de données, plus il faut de l'expertise dans le domaine pour les interpréter au mieux. Ma plus-value dans ce domaine de la *data science* est ma formation de physicien, grâce à laquelle j'apporte des connaissances scientifiques qui permettent une analyse préalable des informations recueillies.

Avez-vous tout de suite décidé de poster vos prédictions sur Twitter? Vous couriez tout de même le risque de vous ridiculiser...

J'ai fait quelques essais avant de les montrer, et j'utilise d'ailleurs mon compte privé et non professionnel pour cette raison. Mais après chaque votation j'ai de plus en plus de followers, ça motive à continuer ces publications.

AUSSI POUR LES AMATEURS DE GRIMPE

Sans le savoir, des Fribourgeois amateurs de grimpe en intérieur ont peut-être déjà bénéficié d'un modèle de prédiction développé par Gradiom. «Au niveau local, je travaille avec le [site grimper.ch](http://site.grimper.ch). Selon la météo, le jour de la semaine ou la période de l'année, je prédis le nombre d'entrées que les salles comptabiliseront dans les trois à cinq jours, ce qui permet de planifier le nombre de moniteurs et de personnel nécessaires. Dans ce type de projets, une grosse partie du temps est passée à obtenir des données et à les traduire dans un langage mathématique et informatique. Mais la partie *machine learning* de transformation des données est ensuite relativement simple et rapide», explique Sébastien Perseguers.

Indirectement, les connaissances acquises avec ces outils d'apprentissage lui ont servi pour d'autres projets professionnels, par exemple dans le traitement de signaux acoustiques. Le *machine learning* lui a ainsi permis d'améliorer les algorithmes devant distinguer le bruit d'un drone, d'une abeille ou des avalanches.

En revanche, son modèle pensé pour les votations devrait être complètement revu s'il voulait l'utiliser pour prédire le vainqueur d'une élection: il faudrait notamment trouver des sources de données fiables ainsi qu'un moyen mathématique de décrire chaque candidat. **TB**