
Analyse du report des voix entre les deux tours des élections législatives anticipées de 2024.

Mémoire en vue de l'obtention du Diplôme d'Etudes Supérieures Universitaires de Data Science
par Alexandre Lainé

1 Introduction

Peu après les élections européennes 2024 auxquelles la liste du Rassemblement National (RN), porté par Jordan Bardella et Marine Le Pen, a obtenu 37 % des suffrages [5]. Le Président de la République a décidé de dissoudre l'Assemblée Nationale et ainsi déclencher des élections législatives anticipées. Après un premier tour record pour le RN, c'est finalement le Nouveau Front Populaire (NFP ou Union de la Gauche, UG) qui est arrivé en tête (182 sièges) après un entre deux tours rythmé par les désistements et les consignes de votes afin de faire barrage au RN dans un maximum de circonscriptions [7].

Ces mouvements de voix sont particulièrement intéressant, et il est possible de se demander si les résultats du premier tour sont suffisant pour prédire le dérouler du second. Basée sur cette interrogation, il s'agirait par conséquent de trouver une règle décrivant les mouvements de voix entre les deux tours. Depuis la publication des résultats, très peu de travaux ont été publiés. La seule approche utilisée à ma connaissance, et ne se basant sur des sondages, est celle de l'apprentissage statistique [1] visant à estimer le taux de report entre les familles politiques. Celui-ci prend alors la forme d'une distribution sur l'ensemble des circonscriptions décrite par une moyenne et une variance particulière. Néanmoins cette approche ne répond pas tout à fait à la question posée, en guise de base à ce travail, il a déjà été décrit une approche d'apprentissage automatique permettant d'estimer le transfert des voix entre les deux tours des élections présidentielles de 2022 sous la forme d'une matrice [6]. Néanmoins, dans notre cas des élections législatives, le problème est plus complexe, car dans chaque circonscription les nuances des candidats du second tour peuvent être différentes. Cela ajoute par conséquent à l'hétérogénéité des bureaux de vote une forte diversité quand au grand nombre de faces à faces possibles.

L'objectif de ce projet est donc d'utiliser cette approche d'apprentissage automatique afin de trouver une matrice représentant le taux de transfert de voix de chaque partie du premier tour vers ceux du second tour. Il est alors nécessaire de passer par plusieurs niveaux de complexité, tout d'abord en regroupant les différentes nuances politiques sous la forme de famille, puis de restreindre l'apprentissage qu'à des cas particuliers de face à face, avant de tenter la phase la plus complexe en tentant l'apprentissage sur l'ensemble des bureaux de votes. Cependant, il est important de notifier quelques précisions. Tout d'abord, il est normalement impossible de trouver le taux réel de report entre deux nuances, car celui-ci est sûrement dépendant de paramètres géographiques et socio-économiques. Ensuite, ce taux de report généralise une règle pour l'ensemble d'une partie et passe donc au dessus des possibles divergences présentes. Enfin, s'agissant d'un sujet actuel, l'objectif n'est ici en aucun cas de "refaire le match" mais seulement d'utiliser des outils issus de l'intelligence artificielle afin d'analyser les données de ces élections en se concentrant sur les mouvements de voix notamment dû à la mise en place d'un "front républicain". Ce terme, apparu pour la première fois à l'occasion des élections législatives de 1956, prend racine dans la défense républicaine qui a eu lieu au début de la troisième république afin de faire barrage aux élans monarchistes.

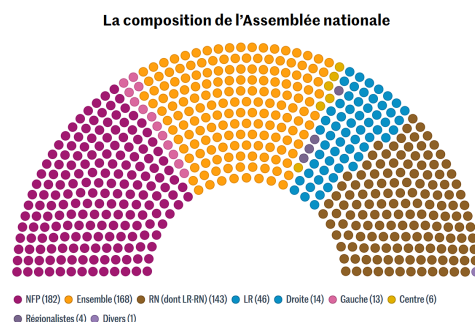


FIGURE 1 – Répartition de l'assemblée nationale suite aux élections législatives anticipées de 2024 [4].

2 Matériel et Méthodes

Jeux de donnés

Résultats des élections législatives pour le premier et second tour pour chaque bureau de vote en France métropolitaine, dans les Territoires d'Outre Mer, mais aussi à l'étranger. Il s'agit respectivement de fichier d'environ 40 et 16 Mo, renseignant pour chaque bureau de vote le nombre de voix obtenu par chaque candidat, ainsi que le nombre d'inscrit, le nombre d'abstention et le nombre de bulletin nul.

Pré-processing des données

De façon général l'objectif de cette phase de préprocessing est de supprimer les colonnes ainsi que les lignes vides, mais aussi les informations qui ne nous seront pas nécessaire pour la suite : genre, nom et prénom du candidat, nom de la commune et du département, ainsi que les bureaux de votes n'ayant pas eu besoin de second tour car ceux-là ne peuvent pas nous renseigner sur le report de voix. Il est important de noter qu'avant toutes étape d'apprentissage, le jeu de donné est coupé en deux parties, la première permettant l'entraînement, et la seconde permettant d'évaluer la précision du modèle sans que cela soit pris en compte pour l'apprentissage. De plus, bien que le tableau ne soit composé que du nombre de voix obtenue par chaque nuance, la première étape dans la fonction d'apprentissage consiste à diviser le nombre de vote obtenue par chaque nuance au sein d'un bureau de vote par le nombre total de vote obtenue dans ce bureau de vote.

Afin d'avoir une progression dans la complexité de la tâche d'apprentissage automatique, on différenciera trois façon de préparer le jeu de donné :

- En regroupant les nuances en familles politiques. L'objectif a été ici de rassembler dans chaque bureau de vote les résultats des nuances présents sous la forme de grand groupes, en se basant sur leur localisation dans l'éventail politiques [8].
- En se focalisant uniquement sur les bureaux de votes présentant des face-à-faces particuliers. Notamment en se focalisant principalement sur ceux ayant généré beaucoup de report dû à la présence du Rassemblement National au second tour.
- En conservant toute la complexité des élections, mais en supprimant tout de même le nombre d'abstention, de vote nul, ou de vote blanc.

Modèle de transfert de voix

Si on considère le pourcentage de voix obtenu par chaque partie dans un bureau de vote comme une distribution discrète notée D , l'hypothèse est qu'il existerait une matrice de transition M permettant de faire le lien entre les distributions du premier et second tour (respectivement D_1 et D_2). Mathématiquement, cela s'exprimerait comme suit :

$$\hat{D}_2 = D_1 \times M \quad (1)$$

L'objectif est donc d'utiliser les résultats du premier et du second tour afin d'apprendre de façon automatique cette matrice M de (n lignes et m colonnes correspondant respectivement au nombre de nuances au premier et au second tour). Ce modèle de transfert de voix sera donc constituer d'une couche d'entrée dont la taille correspond au nombre de nuances politiques présentent au premier

tour dans toute la France, mais aussi d'une couche de sortie, dont la taille correspond au nombre de nuances présentent au second tour. La matrice M correspond par conséquent aux poids reliant la première à la seconde couche.

Divergence de Kullback-Leibler

La divergence de Kullback-Leibler [3] (KL) est une mesure permettant la comparaison entre deux distributions de probabilités discrètes. Celle-ci nous permettra de comparer la distance entre les distributions réelles du second tour (D_2) et celles prédites par notre modèle (\hat{D}_2) selon la formule suivante :

$$KL(D_2, \hat{D}_2) = \sum_{k \in \Omega} D_2 \cdot \log \frac{D_2}{\hat{D}_2} \quad (2)$$

Entraînement et contrôle

Bien qu'il soit possible d'incorporer une fonction de stop anticipé de l'apprentissage, le choix fait ici a été de tenter différents nombre d'itération et de stopper l'apprentissage dès que le résultats de la fonction de coût, pour la partie de validation, se stabilise. La descente de gradient est réalisée par une implémentation pytorch de l'agorithme d'optimisation Adam très efficace dans le cas d'un jeu de donné large et d'une optimisation d'un grand nombre de paramètres [2].

Visualisation et représentations graphiques

Afin de visualiser le jeu de donné et de rendre compte des différents résultats obtenues dans cette pipeline, il a été choisis d'utiliser principalement les bibliothèques matplotlib (version 3.9.0) et seaborn (version 0.13.2) pour leurs simplicités d'utilisation mais aussi de découvrir plotly (version 5.24.0) pour la qualité de ses graphiques interactifs et sa capacité à gérer une grande quantité de données. La plus part des figures créer das la pipeline sont enregistrées de façon automatique en format pdf au sein du dossier "fig" de la pipeline.

Information supplémentaires

L'ensemble des codes de la pipeline sont disponible en libre accès sur la plateforme GitHub en cliquant juste ici. L'ensemble de l'analyse a été développée et écrite pour le mémoire du DESU avec le langage de programmation Python (version 3.11.9) majoritairement sous la forme de notebook appellant des fonctions écrite par mes soins. Les calculs ont été réalisé en local sur mon ordinateur personnel (OMEN by HP Laptop 16-xf0xx) ayant les caratéristiques ci-contre (CPU : AMD Ryzen 9 7940HS, GPU-1 : AMD Radeon 780M, GPU-2 : NVIDIA GeForce RTX 4070).

3 Résultats

Grandes familles politiques

Face-à-face

RN ou UG

RN ou ENS

RN ou LR

Généralisation nationale

4 Discussion

Références

- [1] Victor AMBLARD. *victor-amblard/analyse-legislatives-2024*. HTML. Sept. 2024. URL : <https://github.com/victor-amblard/analyse-legislatives-2024>.
- [2] Diederik P. KINGMA et Jimmy BA. "Adam : A Method for Stochastic Optimization". en. In : arXiv :1412.6980 (jan. 2017). arXiv :1412.6980 [cs]. URL : <http://arxiv.org/abs/1412.6980>.
- [3] S. KULLBACK et R. A. LEIBLER. "On Information and Sufficiency". In : *The Annals of Mathematical Statistics* 22.1 (1951), p. 79-86. ISSN : 0003-4851.
- [4] Le MONDE. "La carte des résultats des législatives 2024 au second tour : la composition de l'Assemblée et le député élu dans votre circonscription". fr. In : (juill. 2024). URL : https://www.lemonde.fr/les-decodeurs/article/2024/07/08/la-carte-des-resultats-des-legislatives-2024-au-second-tour-la-composition-de-l-assemblee-et-le-depute-elu-dans-votre-circonscription_6247510_4355771.html.
- [5] Le MONDE. "Résultats européennes 2024 : les sièges, pays par pays, et la composition du futur Parlement". fr. In : (juin 2024). URL : https://www.lemonde.fr/les-decodeurs/article/2024/06/09/europeennes-2024-les-resultats-en-sieges-pays-par-pays-et-la-composition-du-futur-parlement_6238276_4355770.html.
- [6] Laurent PERRINET. *Elections présidentielles 2022 : estimation du transfert de voix*. en. Juin 2022. URL : <https://laurentperrinet.github.io/sciblog/posts/2022-06-08-transfert-de-voix.html>.
- [7] WIKIPÉDIA. *Élections législatives françaises de 2024*. fr. Page Version ID : 218371077. Sept. 2024. URL : https://fr.wikipedia.org/w/index.php?title=%C3%89lections_1%C3%A9gislatives_fran%C3%A7aises_de_2024&oldid=218371077.
- [8] WIKIPÉDIA. *Liste de partis politiques en France*. fr. Page Version ID : 218218855. Août 2024. URL : https://fr.wikipedia.org/w/index.php?title=Liste_de_partis_politiques_en_France&oldid=218218855.