

## APPEL A PROJETS RECHERCHE 2023

FEDERATION DE RECHERCHE AGORANTIC  
«CULTURE, PATRIMOINES, SOCIÉTÉS NUMÉRIQUES »

### Attention :

- Ne pas dépasser 5 pages
- Transmettre le fichier au format PDF intitulé : ACRONYME-AAP-blc-Agorantic-2023
- Envoyer le fichier à [agorantic@univ-avignon.fr](mailto:agorantic@univ-avignon.fr) **avant le 3 janvier 2023**.

Titre	Analyse des Votes par Transitions
Acronyme	A-Vo-T
Nom du/des porteur(s)	Pierre-Henri Morand
Laboratoires associés	LBNC (AU), INT (CNRS-AMU), ERIC (Lyon II)
Budget demandé	4 730
Résumé Max. 1 000 caractères espaces compris	Le projet A-Vo-T vise à explorer la pertinence d'outils d'apprentissage automatique dans sa capacité à analyser et prédire le comportement de report de vote entre les deux tours d'un scrutin uninominal à 2 tours. En mobilisant les données ouvertes des résultats électoraux des élections présidentielles françaises à l'échelle des bureaux de vote, il permettra d'illustrer le potentiel de ces outils d'analyse des données dans le champ des SHS, de les comparer aux outils statistiques plus traditionnels et d'en identifier les éventuelles limites. Regroupant des chercheurs en statistique et machine learning, neuroscience computationnelle, science politique et économie-économétrie, il s'inscrit dans une démarche exploratoire résolument transdisciplinaire.

### 1. Contexte, positionnement, objectif(s)

Afin d'analyser les résultats des élections, par exemple la dernière élection présidentielle de 2022 en France, et de mieux comprendre la dynamique des choix de vote entre les différents groupes de population, ce projet de recherche vise à mobiliser des outils d'[apprentissage automatique](#) pour inférer des structures à première vue cachées dans la masse des données et de comparer les performances obtenues aux outils plus traditionnels et de confronter les résultats aux connaissances établies sur les comportements électoraux.

Le projet mobilise pour ce faire des chercheurs de 4 disciplines :

- Stéphane Chrétien est professeur de statistique et de machine learning à l'Université de Lyon II.
- Pierre-Henri Morand est économiste, spécialisé dans l'analyse des données ouvertes, professeur au sein du LBNC, Avignon Université.
- Laurent Perrinet est directeur de recherche CNRS, membre de l'INT (CNRS-AMU), spécialiste de neuroscience computationnelle
- Jessica Sainty est politiste, spécialiste de sociologie des comportements politiques et des espaces

politiques locaux. Elle est enseignante-chercheuse en science politique à Avignon Université, membre du LBNC.

Inspiré par un travail exploratoire mené par Laurent Perinnet suite à un [article du Monde](#) mobilisant des données électorales à l'issue du second tour du dernier scrutin présidentiel, le projet A-VO-T vise à explorer la question suivante : peut-on extraire des données brutes des résultats électoraux accessibles en open data à l'échelle la plus désagrégée possible (i.e. à l'échelle des quelques 69 682 bureaux de vote qui archivent les résultats de chaque tour de scrutin des différents scrutins), une estimation du report de voix entre les choix de vote au premier tour et ceux qui sont effectués au deuxième tour. Ces estimations sont-elles robustes et offrent-elles des performances meilleures que les méthodes traditionnelles d'estimation des reports de vote ? Ces prédictions sont-elles par ailleurs conformes aux résultats d'études plus qualitatives menées à l'échelle des bureaux de vote ? Permettent-elles l'identification de catégories pertinentes de comportements électoraux ?

Afin d'estimer la pertinence de l'approche exploratoire conduite, une confrontation avec une approche plus qualitative, qui permet à l'équipe des politistes du LBNC de suivre, à l'échelle de bureaux de votes avignonnais, sur le temps long, le comportement électoral individuel, sera menée. Cela constituera méthodologiquement un point de rencontre particulièrement riche entre approches par la fouille de données massives et méthodes traditionnelles des sciences sociales.

## 2. Problématique, questionnement scientifique, axe(s) de la FR concerné(s)

A-VO-T s'inscrit à l'intersection de deux axes de la Fédération de recherche : AXE 1 Méthodologies et interdisciplinarité et AXE 3 Politique(s), transparence et éthique.

Il vise à expérimenter l'usage des sciences de la donnée dans les SHS et d'identifier le potentiel d'innovation méthodologique que l'apprentissage automatique sur des jeux de données ouvertes massives offre.

En se positionnant sur un traitement d'un jeu de données et des structures de choix « simples », A-VO-T permet également d'analyser finement les apports de ces méthodes par comparaison aux méthodes statistiques plus traditionnelles, d'en valider la pertinence et d'en cerner les limitations.

## 3. Méthodologie

Parmi les outils mathématiques de l'apprentissage automatique, A-VO-T se propose de mobiliser une modélisation de transition des probabilités entre les préférences au premier tour et au second tour.

Cette théorie va permettre d'exprimer le fait que les résultats tels qu'ils sont obtenus peuvent présenter une variabilité mais que celle-ci réelle résulte de préférences de chaque individu dans la population votante. En particulier, on peut considérer que chaque individu va avoir une préférence, graduée entre 0=0% et 1=100% pour chacun des choix (candidats, nul, blanc, abstention) au premier et second tour. Ainsi, leurs comportements électoraux correspondent à la réalisation de ces préférences.

Bien sûr, le vote reste secret et on n'a pas accès au vote de chaque individu et encore moins à ses préférences. Mais comme chaque bureau de vote présente des variabilités liées au contexte local et qui fait que cette population locale a une préférence pour certains choix plutôt que d'autres, on peut considérer chaque bureau de vote comme une population individuelle pour lequel nous allons essayer de prédire les résultats du vote au deuxième tour. En exploitant les irrégularités locales, bureau de vote par bureau de vote, nous allons pouvoir prédire (le mieux possible) le report des

votes individuels (de chaque individu tel qu'il passe d'un vote à un autre).

Grâce à l'ouverture des données sur <https://www.data.gouv.fr/>, on accède librement aux résultats des premier et second tours des scrutins présidentiels. Il est important de noter que les résultats donnés sont indiqués à l'échelle de chacun des bureaux de vote.

Afin de mobiliser une approche sous la forme d'un modèle de transition de probabilités, nous devons faire deux hypothèses que nous allons expliciter mathématiquement :

Tout d'abord, nous allons estimer que pour chaque individu, il existe une préférence pour les choix du premier tour ainsi que pour les choix du deuxième tour.

Le vote de chaque individu se modélise ainsi mathématiquement par un processus de Bernoulli relatif à ces préférences : on peut écrire pour chaque individu les probabilités de vote au premier et au second tour du scrutin (chacune de ces valeurs étant comprises entre 0 et 1 représentant un biais de probabilité pour chacune des alternatives).

Avec une telle modélisation, on peut prédire les résultats du vote, car les préférences de chaque individu pour tel ou tel choix doit se révéler nécessairement au niveau de la population totale. Plus précisément, le théorème central limite indique que la moyenne observée (c'est-à-dire les moyennes de vote observées pour les différents choix) tend vers ces probabilités avec une précision (inverse de la variance) qui augmente linéairement avec le nombre d'observations.

Une deuxième hypothèse que nous allons faire est de considérer la transition entre les préférences qui sont faites au premier tour et celles qui sont faites au second tour.

En effet, les préférences peuvent évoluer avec chaque individu mais que la transition est homogène au sein de la population (par exemple "une personne qui choisit de s'abstenir s'abstiendra"). C'est une hypothèse a priori grossière, mais assez générale pour refléter les tendances au niveau de la population globale, et qui permet d'être testée.

Cette hypothèse est basée sur la modélisation d'une séquence de deux événements aléatoires comme dans un processus de Markov. En particulier, nous allons formaliser cette hypothèse en faisant l'hypothèse de l'existence d'une matrice de transition qui permet de prédire la préférence d'un individu au second tour à partir de ses préférences au premier tour. D'une certaine façon, cette matrice de transition décrit exactement les affinités de chacun des individus pour les différents choix de vote au niveau de la population globale. C'est donc un indicateur quantitatif des reports de vote qui vont être effectivement effectués entre les deux tours.

#### 4. Résultats attendus et caractère innovant de la recherche

Validation croisée de la méthode initialement proposée par L. Perrinet dans son approche exploratoire.

Confrontation de la performance du modèle d'apprentissage automatique par rapport aux outils statistiques traditionnels communément utilisés dans le champ de l'analyse du vote (économie politique et science politique).

Proposition de catégorisation des comportements électoraux et couplage avec les données socio-démographiques liés aux territoires des bureaux de vote.

#### 5. Dimension interdisciplinaire (champs disciplinaires associés) et cohérence par rapport à la thématique « Culture, Patrimoines, Sociétés Numériques »

Le projet s'inscrit ainsi dans une approche résolument pluri-disciplinaire. Il confronte les approches économétriques standards dans l'analyse des votes aux techniques nouvelles issues de l'apprentissage automatique. Il ouvre également un dialogue entre les sciences de la donnée et les analyses traditionnelles du comportement électoral mené en science politique. Il croise enfin trois

champs historiquement travaillés au sein de la Fédération de Recherche Agorantic : comportement électoral à l'échelle du bureau de vote, données ouvertes (ici, données électorales) et apport de l'apprentissage automatique dans le domaine des SHS.

## 6. Partenariats extérieurs envisagés

Le projet est construit en partenariat avec deux structures de recherche qui n'ont jamais participé à la FR Agorantic tout en étant membre des partenaires institutionnels de cette dernière.

**Laurent Perrinet** est directeur de recherche au CNRS, membre de l'Institut des Neurosciences de la Timone (INT, UMR CNRS-Aix Marseille Université). Il est spécialiste en neurosciences computationnelles, se spécialisant sur les mécanismes de prise de décision et leur lien avec des mécanismes d'inférence statistique.

**Stéphane Chrétien** est quant à lui professeur de statistiques et machine learning à l'UFR d'Anthropologie, de sociologie et de science politique et l'Université de Lyon II et chercheur à ERIC (Créé en 1995, le laboratoire ERIC (Unité de Recherche des Universités Lyon 2 et Lyon 1) développe des recherches théoriques et appliquées dans les domaines de la science des données et de l'informatique décisionnelle. Elles visent à valoriser les grandes bases de données complexes, notamment dans les domaines des lettres, langues, sciences humaines et sociales (LLSHS)).

Budget (€)*		
	Brève description	Montant
Missions		
Consommables, petits matériels**		
Organisation de réunions	Déplacements Lyon – Avignon – Marseille du consortium + nuitées	1250
Stages***, <b>vacations</b> , prestations	Stage de M2 – 6 mois	3480
Budget total		4730
Co financements le cas échéant		
Budget demandé à Agorantic		4730
Recettes extérieures		

\* Veuillez modifier les catégories de dépenses si besoin – ajoutez/supprimez des lignes à votre convenance

\*\* Petit matériel ne dépassant pas les 600€

\*\*\*Gratification de stage obligatoire au-delà de 2 mois – prévoir environ 580€ par mois

**NB : si le projet a déjà fait l'objet d'un financement lors d'un précédent AAP : justifier la nouvelle demande et présenter les évolutions du projet.**