

UNIVERSITE PARIS 1 PANTHEON - SORBONNE
2020-2021



Projet ENL :
Prédiction de l'intention de vote
du candidat DSK

Projet réalisé par :
MAHDI DJAMA Djama
SOTO Andrès
BRIDA Kevins

Table of Contents

1	Introduction.....	3
2	Présentation de la base de données.....	4
3	Analyse exploratoire.....	5
3.1	Données manquantes.....	5
3.2	Regroupement des variables.....	6
3.3	Intention de vote de DSK.....	6
4	Modélisation.....	7
4.1	Régressions logistiques.....	7
	Variable de contextes.....	7
	Variables d'opinions.....	10
5	Réseau bayésien.....	12
5.1	L'algorithme IAMB.....	12
5.2	Algorithme HC.....	15
6	Conclusion.....	17

1 Introduction

De nos jours, les systèmes électoraux sont massivement utilisés pour désigner des élus qui exerceront un mandat en tant que représentant du peuple pour gouverner un pays, une ville, voir aussi un village.

Dans le cadre d'élections, le système électoral de prédilection est celui dit au scrutin majoritaire. Le peuple vote pour plusieurs candidats et celui qui récolte le plus de voix est élu. Pour se différencier entre eux, les candidats organisent des campagnes électorales visant à faire la promotion de leurs programmes électoraux. Il n'est pas rare de déceler des tendances de votes en fonction du profil des individus. Par exemple en France, l'électeur moyen du Front national est un jeune homme ouvrier ou employé n'ayant pas ou peu fait d'études.

Le but de notre projet sera d'essayer de prédire au mieux la probabilité de vote d'un candidat à une élection en fonction du profil des électeurs par le biais de plusieurs algorithmes statistiques.

Pour ce faire, nous disposons d'une base de données nous informant sur les intentions de vote des 16 candidats à la présidentielle Française de 2007 au premier tour, en plus d'informations concernant le profil des électeurs ayant émis ces intentions de vote. Le candidat que nous choisissons pour effectuer ce projet est Dominique-Strauss-Kahn (DSK), l'ancien directeur du FMI.

Dans un premier temps nous commencerons par une analyse exploratoire des données visant à présenter brièvement la base de données et toutes les manipulations préalables nécessaire pour la mise en place d'une bonne modélisation.

Ensuite, nous réaliserons plusieurs régressions logistiques visant à prédire de la manière la plus efficace possible la probabilité de vote pour le candidat DSK en fonction des variables d'opinion et de contexte des électeurs. Cette régression nous servira à sélectionner les variables pertinentes. Enfin, nous finirons par un réseau bayésien qui nous permettra de prédire la probabilité de vote pour DSK.

2 Présentation de la base de données

La base de données mise à notre disposition est celle d'une étude faite en 2006 en France se dénommant « Le Baromètre Politique Français 2006 ». Ce baromètre représente un instrument d'étude de la conjoncture politique et de la manière dont les opinions, jugements et comportements politiques se structurent à l'approche des échéances électorales de 2007.

Cette base contient 5600 observations (donc 5600 personnes interrogées) pour 236 variables contenant les intentions de votes des 16 candidats de la présidentielle et d'innombrable variable sur les caractéristiques des électeurs. Une grande partie de ces variables sont de type « catégoriques » c'est-à-dire qu'il ne s'agit pas de variables continues, mais de variables discrètes prenant différentes modalités.

Compte tenu du nombre important de variables, nous décidons de réaliser une sélection préalable visant à réduire considérablement ce nombre pour notre étude. Ainsi, pour la modélisation, nous n'utiliserons que 9 variables explicatives de type « contexte » nous informant entre autres sur le niveau du diplôme, la tranche d'âge, le sexe, la profession, la situation professionnelle, la région, la religion et la situation familiale.

Aussi, nous analyserons 11 variables explicatives de type « opinion » nous informant sur l'opinion des Français sur certains sujets de société, comme leur avis sur les immigrés, le chômage, la délinquance, l'homosexualité...

Enfin, la variable à expliquer sera celle portant sur les intentions de votes du candidat DSK pour la présidentielle de 2007.

3 Analyse exploratoire

3.1 Données manquantes

Afin d'exploiter au mieux notre base de données pour notre partie modélisation, il est primordial de vérifier si les variables que nous avons sélectionnées possèdent des données manquantes.

Parmi nos 20 variables explicatives, seules 3 de ces variables possédaient des données manquantes. Ces données manquantes étaient toutes présentes en très petite quantité, entre 6 et 203 données manquantes pour chacune de ces variables. Il s'agit de 3 variables d'opinion, Q37 (Informations-TV), Q38 (Journal télévisé régulièrement regardé) et Q12B (Problèmes les plus importants pour la France). Nous décidons de les remplacer à l'aide de l'algorithme du KNN plutôt que de les supprimer de la base ou de les remplacer par leur modalité la plus fréquente.

L'algorithme du KNN fait partie de l'apprentissage supervisé, il est utile à la fois pour des problèmes de classification ou de régression. Cette méthode a pour but de classer des points cibles (classe méconnue) en fonction de leurs distances par rapport à des points constituant un échantillon d'apprentissage (c'est-à-dire dont la classe est connue à priori).

Nous l'utiliserons ici, car celui-ci est efficace sur les variables catégoriques. L'utilisation de cet algorithme nous permettra de remplacer ces données de manière homogène.

Pour ce qui est du nombre de voisins, nous laissant la valeur par défaut, c'est-à-dire 5, car le nombre de données manquantes est assez minime, nous ne voyons donc pas d'utilité à réaliser une boucle faisant varier ce chiffre pour trouver le nombre de voisins optimal.

Aussi, nous avons remarqué que certaines variables que ce soit d'opinion ou de contexte, possèdent une modalité « NSP ». La modalité « NSP » (ne se prononce pas) est une modalité choisie par certains électeurs lorsqu'ils ne préfèrent pas répondre à la question. Elles sont présentes en plus grande

quantité pour les variables de type « opinion » sur des sujets sensibles comme l'homosexualité, la peine de mort, la délinquance...

Nous avons décidé de recoder cette modalité en donnée manquante pour que l'algorithme des KNN puisse la répartir de manière homogène entre toutes les autres modalités des variables concernées.

Enfin, notre variable cible l'intention de vote pour DSK ne possède aucune donnée manquante.

3.2 Regroupement des variables

Plusieurs de nos variables explicatives et notre variable à expliquer possèdent un nombre assez conséquent de modalités. Dans le cadre d'une modélisation, il est à notre sens, plus pertinent de réaliser un regroupement pour les variables avec des modalités assez proches entre elles.

C'est pourquoi nous décidons de recoder une partie de nos variables explicatives, l'ensemble des modifications effectuées sont présente dans l'annexe 1.

3.3 Intention de vote de DSK

L'objectif du projet est d'essayer de prédire au mieux une intention de vote positive pour le candidat DSK. Sur les 5649 observations de la base, 1330 individus vont probablement voter pour DSK en 2007, soit 23.5% de l'échantillon, contre 4320 qui ne compte probablement pas le faire. Il s'agit d'un pourcentage plutôt acceptable pour un premier tour compte tenu du fait du nombre élevé de candidats à l'élection.

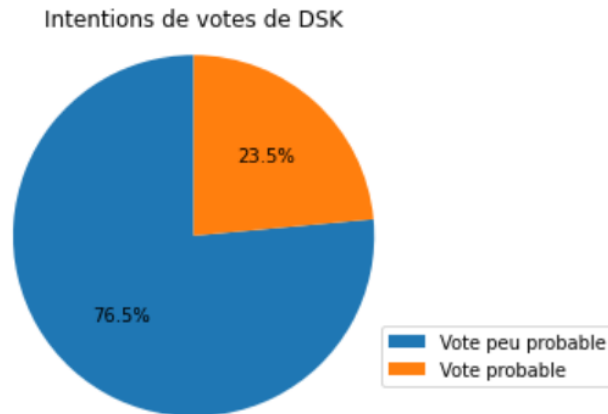


Figure 1 : Diagramme sur les intentions de vote du candidat DSK

4 Modélisation

Dans cette section, nous discuterons de tout ce qui touchera à la modélisation et de comment nous sélectionnerons les meilleures variables selon plusieurs critères.

4.1 Régressions logistiques

Ici, nous réaliserons des régressions logistiques pour sélectionner les variables les plus pertinentes parmi celle d'opinions ou de contextes. Que ce soit pour les variables de contextes ou celles d'opinions, nous décidons de partir sur un échantillon de test de 25 % pour la prédiction et 75% pour l'entraînement des modèles.

Variable de contextes

Afin de sélectionner les variables de contextes les plus pertinentes pour notre futur réseau bayésien, nous décidons de réaliser 512 régressions logistiques avec nos 9 variables de contexte, afin de trouver la combinaison optimale de variables.

Les métriques que nous utiliserons pour sélectionner les modèles pertinents seront l'AIC et le BIC dans un premier temps. Ces deux critères sont utilisés en général pour des prévisions de séries chronologiques, des régressions multiples et des régressions logistiques.

Une fois que nous aurons trouvé le meilleur modèle selon le BIC et l'AIC, nous comparerons leurs AUC (c'est une mesure de la probabilité pour que le modèle classe mieux une probabilité de vote probable dans sa bonne catégorie contrairement à une probabilité peu probable) pour sélectionner le meilleur des deux.

Modèle	AIC	BIC	AUC
Vote ~ RCRS2+RCRS13+RAGE+SEXE+GR+RRS8+RCRS15	4572.384	4623.197	0.5838999
Vote ~ RCRS2+RAGE+SEXE+RCRS15	4578.141	4609.899	0.5780359

Tableau 1 : Les deux modèles qui minimisent le BIC et le AIC

Le tableau ci-dessus nous informe sur les 2 modèles que nous avons sélectionnés grâce au critère du BIC et de l'AIC, avec leurs AUC respectifs.

Ainsi, le premier modèle en bleu est celui qui minimise l'AIC, il possède 7 variables explicatives. Le second en rouge est celui qui minimise le BIC, il possède 4 variables explicatives de contexte. L'AUC du modèle qui minimise l'AIC est de 0.583, contre 0.578 pour celui qui minimise le BIC.

Techniquement, le modèle qui minimise l'AIC est celui qui classifie le mieux notre variable à expliquer. Cependant celui-ci rajoute 3 variables en plus pour un ajout sur l'AUC de quelques centièmes seulement, ce qui semble négligeable. Nous décidons donc de partir sur le modèle qui minimise le BIC, car celui-ci choisit les 4 variables de contexte les plus pertinentes.

Pour affiner notre sélection, nous décidons de tester toutes les interactions possibles entre les 4 variables sélectionnées.

Pour ce faire nous utiliserons 3 méthodes de sélections de variables en fonction de l'AIC et du BIC et nous les départagerons avec le MSE et le RMSE.

- La sélection forward : On cherche d'abord la variable qui explique le mieux y au sens du AIC ou du BIC (minimum) , puis on cherche celle qui, ajoutée à la première, minimise le plus le BIC ou l'AIC, etc.

- La sélection backward : On part du modèle utilisant les p variables explicatives et on cherche, parmi les p variables, celle qui peut être supprimée en occasionnant la plus forte croissance du critère. Cette variable étant supprimée, on itère le processus tant que le BIC ou le AIC n'augmente pas.

- La sélection stepwise : Partant d'un modèle donné, on opère une sélection d'une nouvelle variable (comme avec une méthode ascendante), puis on cherche si on peut éliminer une des variables du modèle (comme pour une méthode descendante) et ainsi de suite.

Ci-dessous se trouvent les résultats que nous avons obtenus :

	Modèle	BIC	RMSE	MSE
Stepwise	RCRS2 + RAGE + SEXE + RCRS15	4609.899	0.4176956	0.1744696
Forward	RCRS2 + RAGE + SEXE + RCRS15	4609.899	0.4176956	0.1744696
Backward	RCRS2 + RAGE + SEXE + RCRS15	4609.899	0.4176956	0.1744696

Tableau 2 : Méthodes de sélection par critère du BIC

Le tableau ci-dessus nous informe sur les résultats des différentes méthodes de sélection qui minimise le critère du BIC. Ainsi, d'après ce tableau, peu importe la méthode de sélection utilisée, le résultat final reste le même, c'est-à-dire notre modèle à 4 variables sans interactions.

	Modèle	AIC	RMSE	MSE
Stepwise	Vote ~ RCRS2 + RAGE + SEXE + RCRS15 + RAGE:RCRS15 + RAGE:SEXE + SEXE:RCRS15 + RCRS2:RAGE	4562.366	0.4167518	0.173682
Forward	Vote ~ RCRS2 + RAGE + SEXE + RCRS15 + RAGE:RCRS15 + RAGE:SEXE + SEXE:RCRS15 + RCRS2:RAGE	4562.366	0.4167518	0.173682
Backward	Vote ~ RCRS2 + RAGE + SEXE + RCRS15	4578.141	0.4176956	0.1744696

Tableau 3 : Méthodes de sélection par critère du AIC

Les méthodes de sélections qui minimisent l'AIC nous donnent des résultats différents. Les méthodes stepwise et forward sont celles qui minimisent le plus l'AIC, celles-ci ajoutent 4 interactions supplémentaires à notre modèle de base. De plus le MSE et le RMSE sont légèrement inférieurs aux résultats obtenus avec le critère du BIC.

L'équation du modèle avec les variables de contextes que nous utiliserons sera donc :

$$Y=B0+B1*RCRS2+B2*RAGE+B3*SEXE+B4*RCRS15 + \\ B5*RAGE*RCRS15+B6*RAGE*SEXE+B7*SEXE*RCRS15+B8*RCRS2 \\ *RAGE$$

Variables d'opinions

Concernant les variables d'opinions, nous utiliserons la même démarche que celles de contextes. Comme ici nous disposons de 11 variables, nous effectuerons dans un premier temps 2048 modèles qui nous serviront à faire le premier tri de nos variables.

Modèle	AIC	BIC	AUC
Vote ~ Q17E+Q12A+Q14+Q17B+Q18+Q38+Q44+Q17D	4476.862	4534.026	0.6430387
Vote ~ Q17E+Q12A+Q38+Q17D	4487.386	4519.144	0.6506490

Tableau 4 : Les deux modèles qui minimisent le BIC et l'AIC

Comme pour les variables de contextes, en rouge nous retrouvant le modèle qui minimise le BIC pour les variables d'opinions et en bleu seulement qui minimise le AIC.

Cette fois-ci, le modèle qui minimise le critère du BIC est celui qui maximise aussi l'AUC, ces 4 variables sont donc celles que nous désignons comme étant les plus pertinentes concernant la probabilité de vote du candidat DSK.

Pour affiner le modèle, nous utiliserons la même méthode que pour les variables de contextes.

	Modèle	BIC	RMSE	MSE
Stepwise	Vote ~ Q17E + Q12A + Q38 + Q17D	4519.144	0.4103802	0.1684119
Forward	Vote ~ Q17E + Q12A + Q38 + Q17D	4519.144	0.4103802	0.1684119
Backward	Vote ~ Q17E + Q12A + Q38 + Q17D	4519.144	0.4103802	0.1684119

Tableau 5 : Méthodes de sélection par critère du BIC

D'après le tableau ci-dessus, pour le critère du BIC, peu importe la méthode de sélection utilisée, le modèle final reste le même.

	Modèle	AIC	RMSE	MSE
Stepwise	Vote ~ Q17E + Q12A + Q38 + Q17D + Q38:Q17D + Q12A:Q38 + Q12A:Q17D	4483.487	0.4104623	0.1684793
Forward	Vote ~ Q17E + Q12A + Q38 + Q17D + Q38:Q17D + Q12A:Q38 + Q12A:Q17D	4562.366	0.4167518	0.173682
Backward	Vote ~ Q17E + Q12A + Q38 + Q17D	4487.386	0.4103802	0.1684119

Tableau 6 : Méthodes de sélection par critère du AIC

Les résultats des méthodes de sélections selon le critère de l'AIC sont différents de celle du BIC pour les méthodes stepwise et forward. Celles-ci suggèrent 3 interactions en plus du modèle initial pour minimiser l'AIC.

Cependant, d'après le RMSE et le MSE sont aux minimums pour les modèles utilisant le BIC et la méthode backward de l'AIC. Il n'est donc pas nécessaire de rajouter des interactions pour les variables d'opinions.

L'équation du modèle avec les variables d'opinion que nous utiliserons sera donc :

$$Y=B0+B1*Q17E+B2*Q17D+B3*Q38+B4*Q12A$$

L'équation finale avec les variables d'opinions et de contexte sera la suivante :

$$Y=B0+B1*Q17E+B2*Q17D+B3*Q38+B4*Q12A+B1*RCRS2+B2*RAGE+B3*SEXE+B4*RCRS15+B5*RAGE*RCRS15+B6*RAGE*SEXE+B7*SEXE*RCRS15+B8*RCRS2*RAGE$$

5 Réseaux bayésiens

Un **réseau bayésien** est un modèle graphique probabiliste représentant un ensemble de variables aléatoires sous la forme d'un graphique. Celui-ci sert à calculer des probabilités conditionnelles, il constitue une base pour les systèmes d'aides à la décision, car il permet de bien représenter les relations causales (diagnostic).

Pour un domaine donné (par exemple financier), on décrit les relations causales entre variables d'intérêt par un **graphe**. Dans ce graphe, les relations de cause à effet entre les variables ne sont pas déterministes, mais **probabilisées**. Ainsi, l'observation d'une cause ou de plusieurs causes n'entraîne pas systématiquement l'effet ou les effets qui en dépendent, mais modifie seulement la probabilité de les observer.

Il s'agit donc d'un algorithme relativement efficace pour répondre à notre problématique. Nous utiliserons 2 fonctions différentes pour le représenter, les fonctions IAMB (Incremental Association Markov Blanket) et HC (Hill-Climbing. La première fonction nous permettra d'intégrer des tests d'indépendances entre les variables, les interactions et notre variable cible et la deuxième d'obtenir des probabilités de vote pour notre candidat DSK pour chacune des variables.

5.1 L'algorithme IAMB

L'algorithme d'IAMB est composé d'une première phase de forward ainsi que d'une seconde phase de backward où les faux positifs sont supprimés. Ces étapes préliminaires simplifient grandement l'identification des voisins. Cela entraîne une réduction significative du nombre des tests d'indépendance conditionnelle, et donc du calcul global de l'algorithme d'apprentissage.

Les options de blacklist introduite sont les suivantes :

	from	to
1	Q17E	RCRS2
2	Q17E	RAGE
3	Q17E	SEXE
4	Q17E	RCRS15
5	Q12A	RCRS2
6	Q12A	RAGE
7	Q12A	SEXE
8	Q12A	RCRS15
9	Q17D	RCRS2
10	Q17D	RAGE
11	Q17D	SEXE
12	Q17D	RCRS15
13	Q38	RCRS2
14	Q38	RAGE
15	Q38	SEXE
16	Q38	RCRS15
17	Vote	Q17E
18	Vote	Q12A
19	Vote	Q17D
20	Vote	Q38
21	Vote	RCRS2
22	Vote	RAGE
23	Vote	SEXE
24	Vote	RCRS15

Tableau 7 : Blacklist réseau bayésien

- Le vote ne peut pas influencer ni le contexte ni l'opinion.
- L'opinion ne peut pas influencer le contexte. Le contexte reste fixe, peu importe l'opinion, ce qui semble logique, la situation professionnelle ne peut être influencée du jour au lendemain par l'opinion.

Les options de whitelist introduite sont les suivantes :

	from	to
1	RAGE	RCRS15
2	RAGE	SEXE
3	SEXE	RCRS15
4	RCRS2	RAGE

Tableau 8 : Whitelist réseau bayésien

Celles-ci correspondent aux interactions significatives que nous avons trouvées suite à la sélection de variables par critère d'AIC pour les variables de contexte.

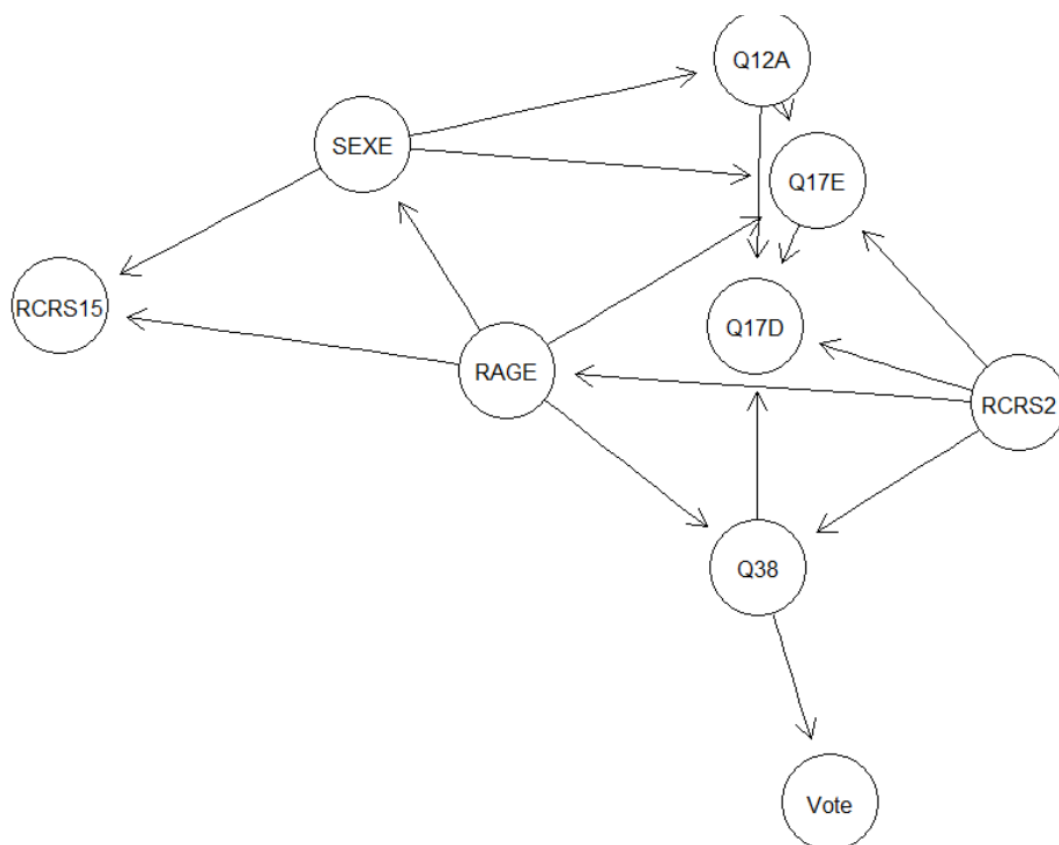


Figure 2 : Réseau bayésien avec fonction IAMB

Ci-dessus se trouve le réseau bayésien utilisant la fonction `iamb` du package « `bnlearn` » de R, avec la whitelist et la blacklist présentées plus haut. Ainsi, d'après le graphique de cet algorithme. Le graphique du réseau bayésien ne contient aucune relation non dirigée indiquant que les variables sont dépendantes entre elles, aucune des variables n'est connectée par un segment sans flèche. Nous en concluons qu'aucune des variables ne se cause mutuellement.

Cependant, il y a 16 relations dirigées indiquant avec une flèche le sens de l'influence d'une variable vers l'autre.

Aussi, il indique que la variable intention de vote pour les candidats DSK est causée par la variable chaînes de TV (Q38). La variable chaîne de TV (Q38) constitue aussi un backdoor entre l'intention de vote, la tranche d'âge (RAGE) et le diplôme (RCRS2).

5.2 Algorithme HC

Les algorithmes de type Hill-climbing (HC), sont des algorithmes fonctionnant sur la maximisation d'un score. Ici, l'algorithme nous permettra d'obtenir des scores de probabilités pour la variable intention de vote de DSK conditionnellement aux variables qui causent ce vote. Nous réaliserons cet algorithme à l'aide de la fonction `hc` présente dans le package `bnlearn`

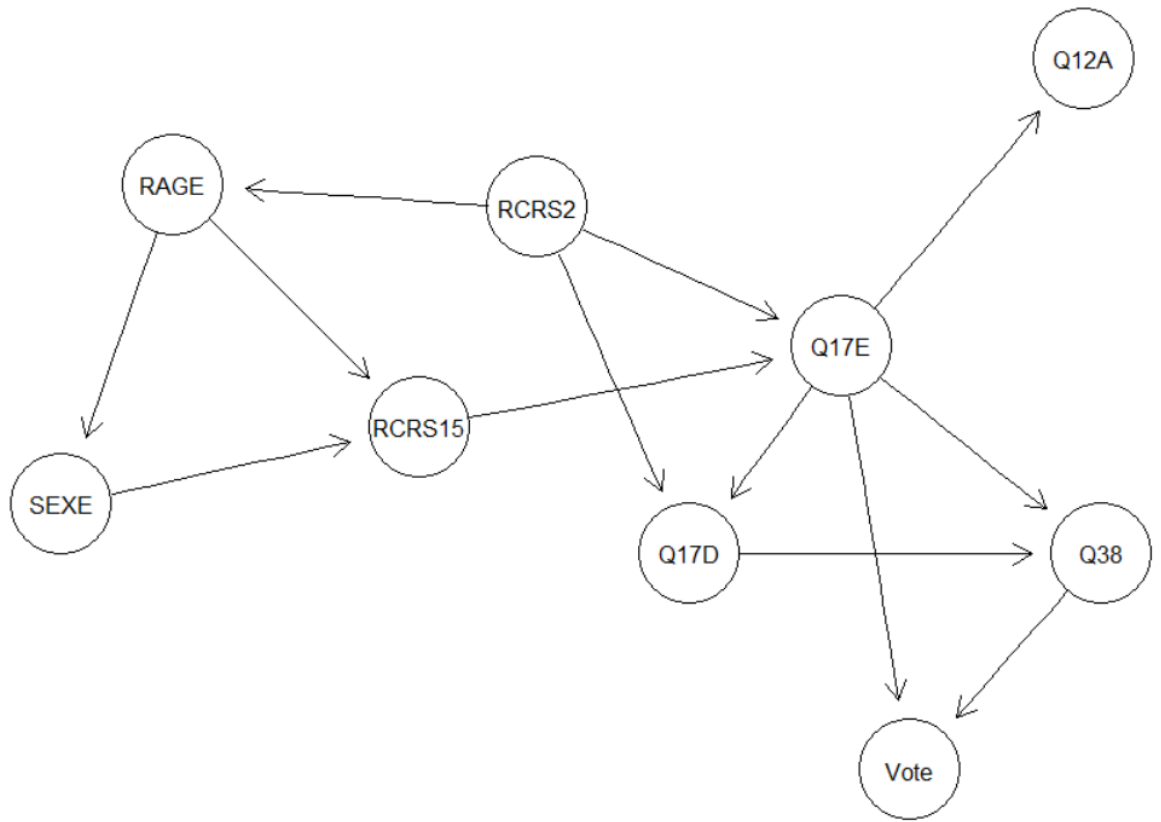


Figure 3 : Réseau bayésien avec fonction HC

Le graphique ci-dessus représente le réseau bayésien que nous avons obtenu à l'aide de la fonction `hc`. Les résultats sont plutôt similaires à ce que nous avons obtenu avec l'algorithme IAMB.

Ici, les variables qui causent l'intention de vote pour DSK sont la chaîne de tv regardée (Q38) et l'opinion sur les immigrants (Q17E). Aussi, les variables Q17E et Q38 constituent des backdoor entre la variable intention

de vote et plusieurs variables. Pour l'opinion sur les immigrés, il s'agit de la religion et du diplôme pour la chaîne de TV regardée, il s'agit de l'opinion sur les immigrés et de la peine de mort.

Grâce à l'algorithme HC, nous avons pu sortir les scores de probabilités pour la variable intention de vote conditionnellement aux variables Q38 et Q17E. Nous les représentons ci-dessous.

Parameters of node Vote (multinomial distribution)			
Conditional probability table:			
, , Q38 = 1			
Q17E			
Vote	1	2	
0	0.8507783	0.7755556	
1	0.1492217	0.2244444	
, , Q38 = 2			
Q17E			
Vote	1	2	
0	0.7693133	0.6420180	
1	0.2306867	0.3579820	
, , Q38 = 3			
Q17E			
Vote	1	2	
0	0.8636364	0.6388889	
1	0.1363636	0.3611111	

Tableau 8 : Probabilités conditionnelles du réseau bayésien en fonction HC

Les probabilités conditionnelles concernant l'intention de vote pour le candidat DSK nous informent que plus les électeurs sont contre le fait de dire qu'il y a trop d'immigration, plus ils seront favorable à voter pour DSK (probabilités conditionnelles plus proches de 1 pour la réponse pas d'accord à Q17E).

6 Conclusion

À l'aide d'une étude sur environ 5600 individus portant sur le baromètre politique français de 2006, il nous était demandé de prédire au mieux les intentions de votes pour le candidat DSK concernant le premier tour des présidentielles Françaises de 2007.

Pour répondre à cette problématique, nous avons préalablement sélectionné 20 variables, 9 variables sur les caractéristiques de contexte des potentiels électeurs et 11 variables sur leurs opinions parmi les 230 disponibles . Afin de créer un réseau bayésien n'utilisant que des variables pertinentes, nous avons réalisé une sélection parmi ces 20 variables à l'aide de régressions logistiques et du BIC/AIC. À l'issue de plusieurs tests statistiques, nous avons réussi à réduire le nombre de variables à 8 avec quelques interactions entre elles.

Enfin, nous avons réalisé deux réseaux bayésiens, le premier pour avoir un premier aperçu sur les relations entre nos variables concernant l'intention de vote pour DSK. Le second nous a été utile pour trouver les scores de probabilité d'intention de vote conditionnellement aux autres variables. Ce réseau bayésien nous a permis de conclure que les électeurs défavorables au fait qu'il y est trop d'immigrés en France sont plus favorables à voter pour DSK.

Annexe

Regroupement des variables

Variable	Libellé variable	Modalité initiale	Modalité après regroupement
Niveau du diplôme	RCRS2	1 = « Sans Dip. , Certif d'études »	1= Sans Dip - BEPC, CAP, SEP
		2= « BEPC, CAP, SEP»	2 = Bac à Dip.Supérieur
		3 = « BAC »	
		4= « BAC +2 »	
		5 = « Dip. Supérieur »	
Parent ou Grand parent étranger	RCRS13	1= A un Parent	1 = A au moins un parent étranger
		2= N'a qu'un G. parent	2 = Pas d'ascendant étranger
		3= Pas d'ascendant étranger	
Situation Familiale	Q48	1=Célibataire,	1 = Célibataire
		2=Marié,	2 = Marié - Couple - Pacsé
		3= Couple sans mariage,	3= Divorcé
		4=Pacsé,	4= Veuf
		5=Divorcé pas en couple,	
		6= Divorcé en couple,	
		7=Veuf pas en couple,	
		8=Veuf en couple,	
Opinion sur le chômage	Q14	1 = Beaucoup augmenté	1 = Augmenté
		2= Un peu augmenté	2 = Stable
		3 = Resté stable	3 = Diminué
		4 =Un peu diminué	
		5=Beaucoup diminué	
Opinion sur la délinquance	Q15	1 = Beaucoup augmenté	1 = Augmenté
		2 = Un peu augmenté	2 = Stable
		3= Resté Stable	3 = Diminué
		4= Un peu diminué	
		5= Beaucoup Diminué	

Opinion sur l'homoséxualité	Q17B	1 = Tout à fait d'accord,	1 = D'accord
		2=Plutôt d'accord,	2 = Pas d'accord
		3= Plutôt pas d'acc.	
		4= Pas du tout d'accord	
Opinion sur la peine de mort	Q17D	1 = Tout à fait d'accord,	1 = D'accord
		2 = Plutôt d'accord,	2 = Pas d'accord
		3= Plutôt pas d'accord,	
		4= Pas du tout d'accord	
Opinion sur les immigrés	Q17E	1 = Tout à fait d'accord,	1 = D'accord
		2 = Plutôt d'accord,	2 = Pas d'accord
		3= Plutôt pas d'accord,	
		4= Pas du tout d'accord, 5=NSP	
Information TV- Nombres de jours regardés	Q37	0= 0 jours ,	1 = De 0 à 3 jours
		1 = 1 jour	2 = De 4 à 7 jours
		2= 2 jours,	
		3=3 jours	
		4= 4 jours,	
		5= 5 jours,	
		6= 6 jours,	
		7= 7jours	
Journal télévisé régulièrement regardé	Q38	1 = 20h TF1,	1 = Ligne éditoriale à droite
		2 = 13h TF1,	2 = ligne éditoriale à gauche
		3 = 20h FR2,	3 = autres
		4 = 13h FR2,	
		5 = 19-20h FR3,	
		6 = Soir 3 FR3,	
		7 = 12-14 FR3,	
		8 = Canal+,	
		9 = Arte Info,	
		10 = 6 min M6,	
		11= 12h50 M6,	
		12 = LCI,	
		13 = I-télé,	
		14 = Autre,	
		15 = Aucun,	

Intentions de votes pour DSK	Q46M	1= Tout à fait probable,	1 = Probable
		2= Plutôt probable,	0 = Pas probable
		3=Plutôt pas probable,	
		4=Pas probable,	
		5=Ne connaît pas,	