

Comparaison des modèles de prédiction électorale: Québec-ROC vs Six Régions

CLESSN

Invalid Date

Table des matières

1	Résumé exécutif	2
2	Introduction	2
3	Méthodologie comparative	3
3.1	Données et environnement de test	3
3.2	Métriques d'évaluation	3
4	Comparaison des performances globales	3
4.1	Métriques d'accuracy et de recall	3
4.2	Performance par parti politique	3
4.3	Performance en Top-2 Accuracy par parti	4
5	Comparaison des performances régionales	5
5.1	Performance par région	5
5.2	Analyse détaillée du Reste du Canada	5
6	Analyse de robustesse statistique	6
6.1	Comparaison des résultats bootstrap	6
7	Avantages et inconvénients de chaque approche	6
7.1	Modèle Québec-ROC	6
7.2	Modèle Six Régions	7
8	Recommandation et conclusion	7
8.1	Recommandation	7
8.2	Conclusion	7
9	Annexes	8
9.1	Définition des régions	8
9.2	Partis politiques	8

1 Résumé exécutif

Ce rapport présente une comparaison détaillée entre deux modèles de prédiction électorale développés par le CLESSN:

1. **Modèle Québec-ROC:** Un modèle qui divise le Canada en deux blocs (Québec et Reste du Canada)
2. **Modèle Six Régions:** Un modèle plus granulaire qui décompose le Canada en six régions distinctes (Québec, Ontario, Colombie-Britannique, Prairies, Atlantique, Territoires)

Les résultats de notre analyse comparative révèlent que:

- Le modèle Six Régions offre une meilleure performance globale avec une accuracy de 54,75% contre 54,19% pour le modèle Québec-ROC
- La performance est améliorée dans 5 des 6 régions, avec des gains particulièrement importants en Colombie-Britannique (+4,15 points) et dans les Territoires (+8,74 points)
- Le modèle Six Régions montre une robustesse statistique nettement supérieure avec un intervalle de confiance à 95% plus élevé (55,94%-63,20% contre 48,45%-54,04%)
- Le F1-score est amélioré pour le CPC, le LPC et le NPD, avec une légère diminution pour le BQ et le GPC

Sur la base de ces résultats, nous recommandons l'adoption du **modèle Six Régions** pour les analyses électorales futures, en raison de sa précision supérieure, de sa robustesse statistique plus élevée et de sa capacité à capturer les dynamiques régionales spécifiques.

2 Introduction

Les modèles de prédiction électorale sont essentiels pour comprendre les dynamiques politiques et anticiper les comportements de vote dans différentes régions. Dans cette optique, nous avons développé et testé deux approches distinctes pour modéliser les intentions de vote au Canada:

1. **Modèle Québec-ROC:** Cette approche divise le pays en deux blocs - le Québec et le Reste du Canada (ROC) - en raison des différences linguistiques, culturelles et politiques significatives entre ces deux régions.
2. **Modèle Six Régions:** Cette approche plus granulaire divise le Canada en six régions distinctes (Québec, Ontario, Colombie-Britannique, Prairies, Atlantique, Territoires) pour mieux capturer les spécificités régionales.

Ce rapport vise à comparer systématiquement ces deux modèles pour déterminer lequel offre les meilleures performances et la plus grande robustesse prédictive.

3 Méthodologie comparative

3.1 Données et environnement de test

Les deux modèles ont été évalués sur des ensembles de validation représentant 30% des données disponibles, issus du même processus de séparation (avec le même seed=42). Les deux modèles ont été entraînés sur les mêmes variables explicatives, mais avec des interactions régionales différentes.

3.2 Métriques d'évaluation

Pour comparer objectivement les deux modèles, nous utilisons plusieurs métriques complémentaires:

- **Accuracy globale:** Proportion générale des prédictions correctes
- **F1-Score par parti:** Moyenne harmonique de la précision et du rappel
- **Top-2 Accuracy:** Proportion des cas où la vraie classe est parmi les deux premières prédictions
- **Performance régionale:** Accuracy dans chaque région géographique
- **Analyse bootstrap:** Estimation de la robustesse et de la variabilité statistique

4 Comparaison des performances globales

4.1 Métriques d'accuracy et de recall

Table 1: Comparaison des métriques globales entre les deux modèles

Métrique	Modèle_Quebec_ROM	Modèle_Six_Régions	Différence	Amélioration_Pct
Accuracy globale	0.5419	0.5475	0.0056	1.03
Recall moyen (macro)	0.4509	0.4527	0.0018	0.40
Accuracy top-2	0.8272	0.8275	0.0003	0.04

Le modèle Six Régions présente une amélioration de 0,56 points de pourcentage en accuracy globale (54,75% contre 54,19%) et de 0,18 points de pourcentage en recall moyen (45,27% contre 45,09%). L'accuracy du premier ou deuxième choix (top-2) est légèrement supérieure dans le modèle Six Régions (82,75% contre 82,72%). Ces améliorations, bien que modestes en apparence, représentent des gains relatifs significatifs pour des modèles de prédiction électorale à cinq partis.

4.2 Performance par parti politique

Table 2: Comparaison des F1-scores par parti politique

Parti	Quebec_ROC	Six_Régions	Différence
BQ	0.6556	0.6548	-0.0008
CPC	0.5289	0.5362	0.0073
LPC	0.4722	0.4853	0.0131
NDP	0.5091	0.5112	0.0021
GPC	0.0519	0.0416	-0.0103
Moyenne	0.4435	0.4458	0.0023

L'analyse par parti révèle des différences importantes dans la performance des deux modèles:

- Le **Parti Libéral (LPC)** voit sa prédiction s'améliorer de 1,31 points avec le modèle Six Régions
- Le **Parti Conservateur (CPC)** gagne 0,73 points
- Le **Nouveau Parti Démocratique (NPD)** gagne 0,21 points
- Le **Bloc Québécois (BQ)** et le **Parti Vert (GPC)** montrent de légères diminutions de performance (-0,08 et -1,03 points respectivement)

En moyenne, le F1-score tous partis confondus est légèrement supérieur dans le modèle Six Régions (+0,23 points).

4.3 Performance en Top-2 Accuracy par parti

Table 3: Comparaison de l'accuracy top-2 par parti politique

Parti	Quebec_ROC	Six_Régions	Différence
BQ	0.9571	0.9596	0.0025
CPC	0.7705	0.7724	0.0019
LPC	0.9024	0.9051	0.0027
NDP	0.6712	0.6602	-0.0110
GPC	0.0437	0.0360	-0.0077

L'analyse de l'accuracy du premier ou deuxième choix (top-2) montre des résultats mixtes:

- Trois partis (BQ, CPC, LPC) voient leur top-2 accuracy légèrement améliorée avec le modèle Six Régions
- Deux partis (NPD, GPC) montrent une légère diminution
- Le BQ maintient une excellente performance dans les deux modèles (>95%)
- Le LPC atteint plus de 90% de top-2 accuracy dans les deux modèles

Ces résultats suggèrent que les deux modèles sont très performants pour identifier le premier ou deuxième choix des électeurs, avec de légères variations selon les partis.

5 Comparaison des performances régionales

5.1 Performance par région

Table 4: Comparaison de l’accuracy par région

Région	Quebec_ROC_Accuracy	Six_Régions_Accuracy	Différence
Québec	0.5348	0.5299	-0.0049
Ontario	0.5487	0.5647	0.0160
Colombie-Britannique	0.4816	0.5231	0.0415
Prairies	0.5799	0.5939	0.0140
Atlantique	0.6010	0.6052	0.0042
Territoires	0.4483	0.5357	0.0874

La décomposition régionale révèle des différences significatives:

- **Améliorations dans 5 régions sur 6:** Le modèle Six Régions surpasse le modèle Québec-ROC dans toutes les régions sauf le Québec
- **Gains majeurs:** Territoires (+8,74 points), Colombie-Britannique (+4,15 points), Ontario (+1,60 points), Prairies (+1,40 points)
- **Gain modeste:** Atlantique (+0,42 points)
- **Légère diminution:** Québec (-0,49 points)

Cette variation indique que le modèle Six Régions permet un meilleur ajustement aux spécificités de la plupart des régions, au prix d’une légère perte de performance au Québec.

5.2 Analyse détaillée du Reste du Canada

Dans le modèle Québec-ROC, le Reste du Canada est traité comme un bloc uniforme:

Table 5: Performance du modèle Québec-ROC par bloc régional

Bloc_Régional	Accuracy	Top2_Accuracy
Québec	0.5348	0.8122
Reste du Canada (ROC)	0.5506	0.8460

Le modèle Six Régions permet d’obtenir une vision plus détaillée des performances au sein du ROC:

Table 6: Performance détaillée du modèle Six Régions dans le ROC

Région	Accuracy
Ontario	0.5647
Colombie-Britannique	0.5231
Prairies	0.5939

Région	Accuracy
Atlantique	0.6052
Territoires	0.5357
Moyenne ROC	0.5645

Cette décomposition révèle des variations importantes au sein du ROC: l'accuracy varie de 52,31% en Colombie-Britannique à 60,52% dans l'Atlantique. La moyenne pondérée de l'accuracy dans les régions hors Québec (56,45%) est supérieure à l'accuracy globale du bloc ROC dans le modèle Québec-ROC (55,06%), ce qui démontre l'intérêt de la modélisation régionale détaillée.

6 Analyse de robustesse statistique

6.1 Comparaison des résultats bootstrap

Table 7: Comparaison des résultats de l'analyse bootstrap

Métrique	Quebec_ROC	Six_Régions
Accuracy moyenne	0.5177	0.5963
Écart-type	0.0138	0.0171
Borne inférieure IC-95%	0.4845	0.5594
Borne supérieure IC-95%	0.5404	0.6320

L'analyse bootstrap révèle des différences majeures en faveur du modèle Six Régions:

- Une accuracy moyenne nettement supérieure (59,63% contre 51,77%, soit +7,86 points)
- Un intervalle de confiance à 95% entièrement plus élevé (55,94%-63,20% contre 48,45%-54,04%)
- Une variabilité légèrement supérieure (écart-type de 0,0171 contre 0,0138)

Ces résultats indiquent que le modèle Six Régions est non seulement plus précis mais aussi significativement plus robuste statistiquement, avec une borne inférieure de confiance supérieure à la borne supérieure du modèle Québec-ROC. La légère augmentation de la variabilité est largement compensée par l'amélioration substantielle de la performance moyenne.

7 Avantages et inconvénients de chaque approche

7.1 Modèle Québec-ROC

Avantages: - Simplicité conceptuelle et computationnelle - Meilleure performance pour le Québec (+0,49 points) - Performance légèrement supérieure pour le Bloc Québécois (+0,08 points en F1-score) - Performance légèrement supérieure pour le GPC (+1,03 points en F1-score)

Inconvénients: - Traite le ROC comme un bloc homogène malgré ses disparités - Performance globale inférieure (-0,56 points d'accuracy) - Performance régionale inférieure dans 5 régions sur 6 - Robustesse statistique significativement plus faible - Moins performant pour les partis nationaux (CPC, LPC)

7.2 Modèle Six Régions

Avantages: - Meilleure accuracy globale (+0,56 points) - Performance supérieure dans 5 régions sur 6 - Robustesse statistique nettement supérieure (IC 95%: 55,94%-63,20%) - Meilleure prédiction pour le CPC (+0,73 points) et le LPC (+1,31 points) - Capacité à capturer les spécificités régionales - Amélioration substantielle en Colombie-Britannique (+4,15 points) et dans les Territoires (+8,74 points)

Inconvénients: - Complexité accrue du modèle - Performance légèrement réduite au Québec (-0,49 points) - Performance légèrement réduite pour le BQ (-0,08 points) et le GPC (-1,03 points) - Variabilité légèrement plus élevée

8 Recommandation et conclusion

8.1 Recommandation

Sur la base de notre analyse comparative approfondie, nous recommandons fortement l'adoption du **modèle Six Régions** pour les futures analyses électorales, pour les raisons suivantes:

1. **Performance supérieure:** Il offre une meilleure accuracy globale (+0,56 points) et des performances régionales supérieures dans 5 régions sur 6
2. **Robustesse statistique nettement supérieure:** Son intervalle de confiance à 95% (55,94%-63,20%) est entièrement supérieur à celui du modèle Québec-ROC
3. **Granularité régionale:** Il permet de capturer avec plus de précision les disparités régionales au sein du Canada
4. **Amélioration majeure pour les partis nationaux:** Meilleure prédiction pour le CPC et le LPC

Bien que le modèle Québec-ROC reste légèrement plus performant au Québec, le modèle Six Régions offre une vision plus nuancée et globalement plus précise du paysage électoral canadien.

8.2 Conclusion

Cette étude comparative démontre qu'une approche régionale plus granulaire améliore significativement la qualité des prédictions électorales au Canada. Le modèle Six Régions, en prenant en compte les spécificités de l'Ontario, de la Colombie-Britannique, des Prairies, de l'Atlantique et des Territoires en plus du Québec, permet de mieux saisir les dynamiques électorales locales et d'obtenir des prédictions plus précises.

L'amélioration substantielle de la robustesse statistique, avec une accuracy bootstrap de 59,63% contre 51,77% pour le modèle Québec-ROC, constitue l'argument le plus convaincant en faveur du

modèle Six Régions. Cette différence de 7,86 points indique que le modèle à six régions est non seulement plus précis en moyenne, mais aussi beaucoup plus stable et fiable.

Ces résultats soulignent l'importance d'adapter la modélisation à la diversité régionale du Canada et ouvrent des perspectives pour des analyses plus fines des comportements électoraux à travers le pays.

9 Annexes

9.1 Définition des régions

- **Québec:** Province de Québec
- **Ontario:** Province d'Ontario
- **Colombie-Britannique:** Province de Colombie-Britannique
- **Prairies:** Alberta, Saskatchewan, Manitoba
- **Atlantique:** Nouveau-Brunswick, Nouvelle-Écosse, Île-du-Prince-Édouard, Terre-Neuve-et-Labrador
- **Territoires:** Yukon, Territoires du Nord-Ouest, Nunavut

9.2 Partis politiques

- **BQ:** Bloc Québécois
- **CPC:** Parti Conservateur du Canada
- **LPC:** Parti Libéral du Canada
- **NDP:** Nouveau Parti Démocratique
- **GPC:** Parti Vert du Canada