**Project Description: Canada Votes**

The goal was to investigate to what extent voting for a particular party depends on structural factors. The federal election results of 2004, 2006, 2008, 2011, and 2015 elections are mapped against key factors from the 2016 Census for each electoral district.

Several machine learning models are used to, and their accuracy is determined. The key factors chosen were (for each electoral district) population, population density, average age, average household size, average total income, percent of immigrants, % of mother-tongue English, French, and Other and poverty rate. The machine learning models used were Random Forest, Nearest Neighbor, and SVM(linear).

Forecast accuracy was 63% for Nearest Neighbor and Random Forest, around 52% for SVM. Tuning of the models did not show any improvements. These results point towards a healthy democratic environment - while structural factors have some influence towards votes, the election program and the popularity of the parties decide the outcome of the elections.

If the goal is to improve the model accuracy, it would be necessary to include party/party leader popularity (which has been excluded as the goal is to investigate the underlying structure) and to look at election-district specific voting behaviour, may be supported by specific polls.

A disadvantage of the model is that it does not fit for very small parties (Green and Independent have been excluded), and that it will not forecast new developments (like the 'orange wave' in 2011 or possible consequences of the rising popularity of the Green party/weakness of the NDP. Especially in a 'first'past-the-post' system these developments can have significant results.

For the above reasons, using just this model for result forecasting will provide inaccurate results and is not recommended. However, just in case somebody is interested - the most realistic result of the 'structural' forecast is Liberal 112 seats, Conservative 136 seats, NDP 45 seats and Bloc 44 seats. I don't think that this is a realistic forecast due to the relative weakness of the Conservative program, the rising popularity of the Green party, and the current weakness of NDP and Bloc.

Accuracy of forecast with different datasets:



Forecast, and results of past elections:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Election | Liberal | Conservative | NDP | Bloc |
| 2004 | 135 | 99 | 19 | 54 |
| 2006 | 103 | 124 | 29 | 51 |
| 2008 | 77 | 143 | 37 | 49 |
| 2011 | 34 | 166 | 103 | 4 |
| 2015 | 184 | 99 | 44 | 10 |
| 2019Fc - NN | 103 | 151 | 34 | 49 |
| 2019Fc - RF | 112 | 136 | 45 | 44 |
| 2019Fc - SVM(lin) | 67 | 196 | 0 | 74 |