

Canadian Federal Election Candidate Analysis

DSE200x - Dave Jordison

Abstract

Dataset: Kaggle “Canadian Federal Election Results (Timeseries)” (subset of 1997 and 2000 election years)

Research Question: Impact of non-political factors contributing to individual candidate electoral success.

Methods: Pre-political occupation with text analysis and wordClouds, impact of incumbency with classification, prediction of electoral success using one-hot encoder decision tree with feature importance.

Key findings: Lawyers and students poor indicator of success, candidate gender unimportant, and incumbent status is the dominant indicator of successful election.

Motivation

Evaluation of factors to influence political success could be critical for persons seeking political office, as well as where parties can deploy resources for maximum effectiveness.

Dataset used:

Canadian Federal Election Results (Timeseries).

<https://www.kaggle.com/smidth80/canadian-federal-election-results-timeseries>

(186MB)

Contains: Election ID, event number, electoral district, ballot sequence, candidate name, occupation, sex, party affiliation, incumbency indicator, and election indicator from 1997 to 2015.

Data Preparation and Cleaning

Data was cleaned for missing rows and NaN values.

Elections exclusive of 1997 and 2000 missing indicator of successful election, no additional datasets were available to supplement missing data.

Decision Tree pre-processing: Subset selection of most frequency occupations with one-hot encoding (nominal categorical data).

Research Question

What are the important aspects of individual Canadian Federal political success, exclusive to party affiliation?

Methods

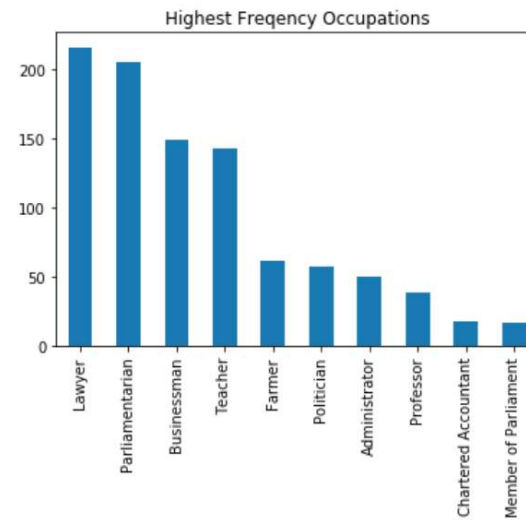
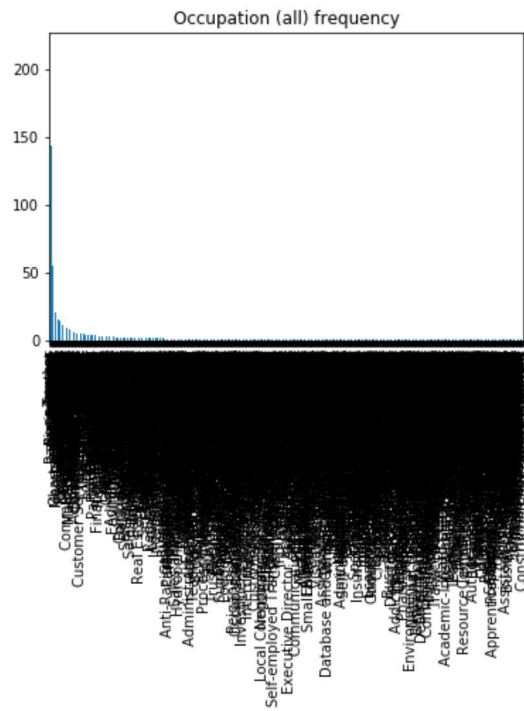
Occupation analyzed via bag of words text methods and wordCloud generation for visual inspection of occupations for both winning and losing candidates.

Incumbency success analyzed via filtering and tabulation of individual candidates across major parties.

Decision tree used to predict individual candidate success, with feature importance. One-hot encoding and dimensionality reduction required for occupations, label encoding required for other categorical variables.

Findings - Occupation

Logarithmic distribution of occupation



Findings - Occupation

Most common occupations for successful federal elected are largely political.

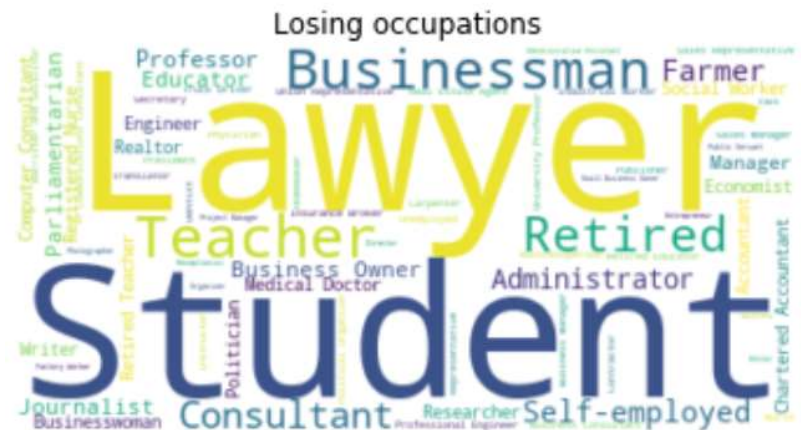
Non-federal politics or business would be good pre-federal occupations for those seeking federal office.



Findings - Occupation

Most common occupations for unsuccessfully elected: Student (unsurprising) & Lawyer (surprising)

Those seeking federal office may wish to avoid these career tracks.



Occupation	Win	Lose

Lawyer	55	181
Parliamentarian	199	32

Findings - Incumbency

Incumbency critical

Very difficult to break into success, relatively easy to retain position

On margin, party resources may be best spent securing existing positions rather than seeking further positions.

Status	Win	Lose
Incumbent	458	67
Not Incumbent	132	2789

Prediction of Electoral Success

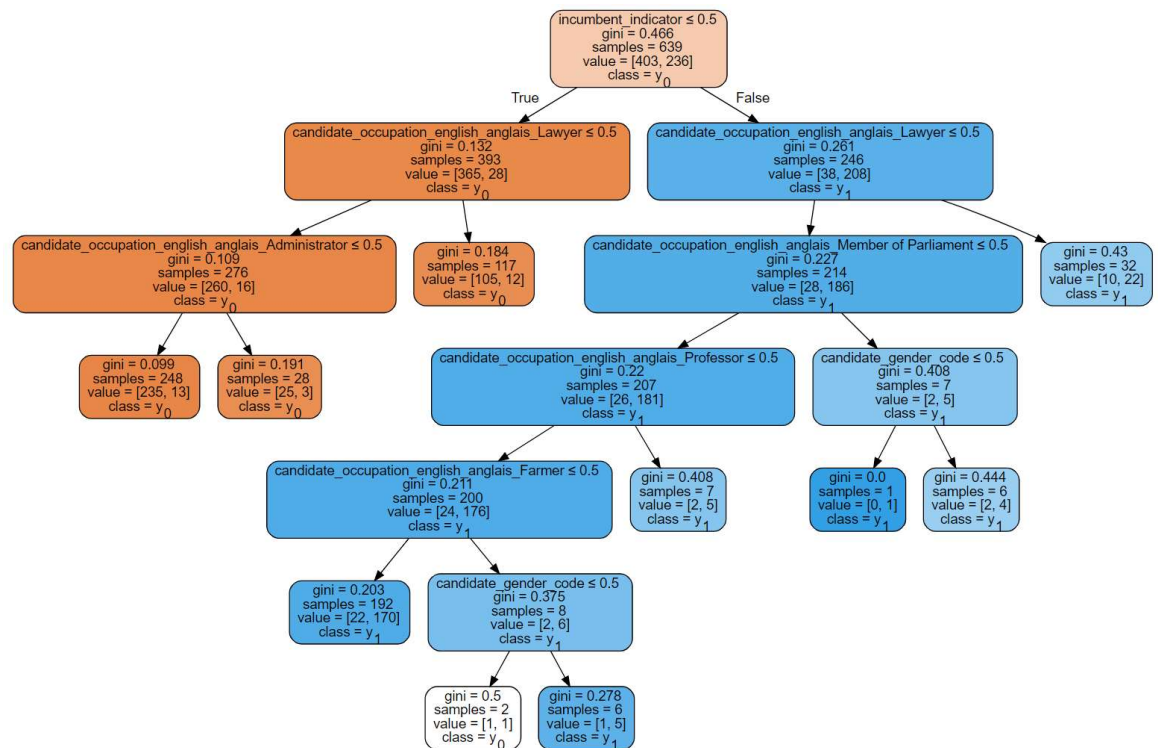
Prediction accuracy 93% with
33% test size

Encoding:

Occupation: (1 = Bool True)

Gender (0 = Female)

Incumbent (1 = Incumbent)



```
accuracy_score(y_true = y_one_hot_test, y_pred = predictions_one_hot)
```

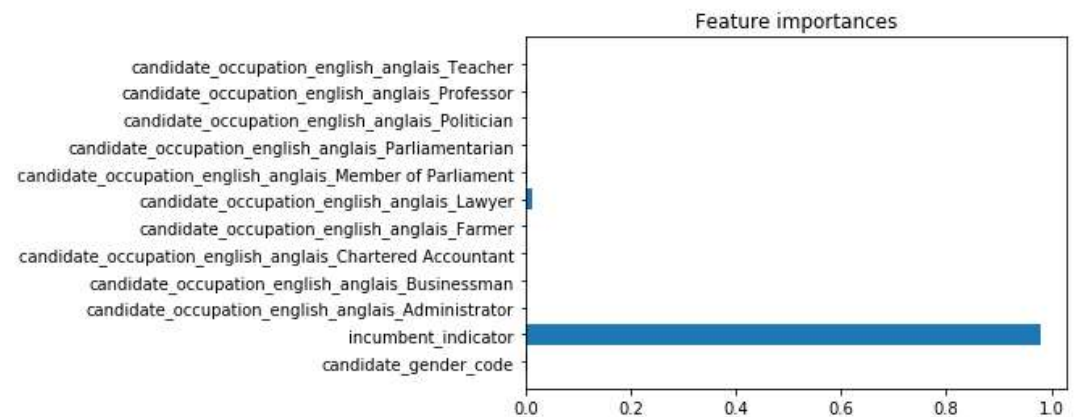
0.926984126984127

Prediction of Election

Incumbency dominant feature

Gender unimportant

Occupation unimportant
(excl: Lawyer)



Limitations

Data considers only Canadian federal election.

Complete data only available for 1997 and 2000 election years, in both events the majority party remained in power. Elections with change of power may not fit model.

No Geodata on ridings available for this period, no visual analysis possible.

Social media in 1997-2000 non-existent, likely that this incumbency-dominant model may no longer be valid and superseded by other factors.

No election data yet available for 2019 Canadian federal election predictions (Q1 2020).

Conclusions

Candidates seeking federal election with backgrounds in politics or business more successful; lawyers and students less successful. Impact of less common careers not known due to small sample size.

Gender not a significant factor to election.

Incumbent status strongly supports success. Parties should use discretionary resources accordingly in alignment with strategic targets.

Acknowledgements

Data collected from Kaggle data source. Work is original and done without outside influence or assistance.

References

No references – all work completed independently.