Group Report

180015455, 180027292, 180029243

Due 19 April 2019

Points to think about:

IDEA: Sales pitch - sell the consulting team, rather than the models themselves

- This should be brief, no more than 4 pages in total.
- There should be a very succinct executive summary at the start selling your method/results.
- What does your approach offer this type of business? Why should they part with a large consulting fee or licence your model? If you can argue a good Return On Investment (ROI) then you're in a good position. Feel free to speculate about costs associated with the different decisions implied. Also bear in mind this document is something of a sales pitch - it should look the part, not some dry analyst's report.
- You should also attempt to give some insight into what is driving the response.
- You have many variables at your disposal and the client would appreciate some insight into what characterises the targets. (Note this is where the client can give your model the "sniff-test" for anything suspicious - 'client-visible' problems).

Executive Summary

- data
- aims
- method
- findings
- applications

Introduction

In the current era of political uncertainty presented by the United Kingdom's leave or remain referendum the nation's MPs have been faced with tough decisions. On the 27^{th} of March 2019, 649 MPs voted on eight alternative Brexit options, all of which failed to pass. This presented the unqiue opportunity to build models aimed at predicting party and constituency representation through the voting profiles of parliamentary members.

The voting profiles of 649 UK parliamentary members were obtained from an online article published by The Guardian (The Guardian, 2019) using Python 3.7 software [@Pyt]. This analysis aims to identify the most reliable models that can use an MP's vote profile to predict both the party they represent as well as the leave/remain percentage for their constituency in the Brexit referendum.

Methods

A series of different models were fitted, their performance elvaluated and the two most reliable models have been presented.

Performance was based on a model's ability to correctly classify the MPs party. For example, Conservative Party members are correctly predicted to be Conservative Party members rather than Labor Party members, and members who are not Conservative Party members should not be predicted to be Conservative Party members.

Analysis and Results

All analyses in the following report were carried out using R 3.5.1 software (R, 2019).

Predicting Party Represenation

When predicting the Party of each MP based on how they voted on the alternative Brexit options, it was found that the Random Forest model made predictions that were closer to the original data than any other model we fitted. Random Forests are classification models that output probabilities of outcomes, in order to support decision making. This Random Forest model can be expected to predict the correct Party of each MP 97.35% of the time. See Table… in Appendix A.XXX for a full set of results obtained from model comparison. It shows that the Random Forest outperforms the prediction of other models built. We are confident that this Random Forest model provides the best predictions possible for these data.

Predicting Leave/Remain Percentage

Conclusions

Appendix

Table 1: Accuracy of each model

Model	Accuracy
Classification Tree	94.70%
Random Forest	97.35%
Logistic Regression	94.04%
Naive Bayes	78.15%
Support Vector Machines	94.04%
Neural Net	93.38%

Table 2: Accuracy of each model

Model	Accuracy of Percentage	Accuracy of Leave/Remain
Classification Tree	111.96	94.70%
Random Forest	109.00	97.35%
Generalized Linear Model	116.10	94.04%
Support Vector Machines	120.27	94.04%
Neural Net	131.67	93.38%

OTHER STUFF

Abtin intro: This analysis takes unique opportunity that has been presented by the current political affairs to a statistical trial. In the current era of political uncertainty presented by the United Kingdom's Brexit referendum, the nation's MPs have been faced with tough decisions; decisive matters that shape it's future. The UK's prime minister, Theresa May, has attempted to come to an agreement with the European Union. This agreement would define the socio-economic dynamic between the United Kingdom and the EU. This provides an interesting data set of each MPs affiliated party, constituency, and their voting record for the eight motions proposed.

Aim one gives useful insight in the alignment and constituency of members association to their party. It is also important to note that many members did not stand with their party in certain motions and presented a power struggle. A change in the dynamic of the parties have been seen to the point of MP's accepting to vote for the Brexit deal if the Prime minister resigns. Aim two provide very important insight in the ability to check, predict and configure the voting patterns (behavior) of each MP. This is with respect to the degree of representativeness to the vote of people whom they represent.

The biggest political parties in the House of Commons are the Conservative Party (party of the Prime Minister, Theresa May) and the Labour Party.

Lei Methods: There are two different tasks of this data mining work, we chose different model schemes based on the attributes of prediction data for each task. For the task one, the model scheme includes Classification Tree, Random Forest, Logistic Regression, Naïve Bayes, Support Vector Machine and Neural Net. For the task two, the model scheme includes the Decision Tree, Random Forest, General Linier Model, Support Vector Machine and Neural Network. The process of model training and optimization follows a standard workflow, the Data-drive workflow (T W Kelsey, 2019), to make sure all work is running in the correct way.

In order to evaluate performance of models and seek the best model for each task, we build a framework about how to evaluate and compare different models. In the unified framework, all models will be trained and tested by the same size dataset and evaluated by the same indicators. For the task one, a classification problem, we hope to find out the best model which could correctly select the right classification and correctly reject the misclassification.

For example, Conservative Party members are correctly predicted to be Conservative Party members rather than Labor Party members, and members who are not Conservative Party members should not be predicted to be Conservative Party members. For the task two, the logic is similar, we predict the percentage of each Constituency and translate these percentage to remain/leave for easier evaluation. When we translated the percentage prediction, we could use the same framework to evaluate and compare different models to find out the best model for the task two. The comparison of pure data is difficult to get the best model quickly. In order to better show the performance of different models, we visualize the results of model evaluation.

Possible plot:

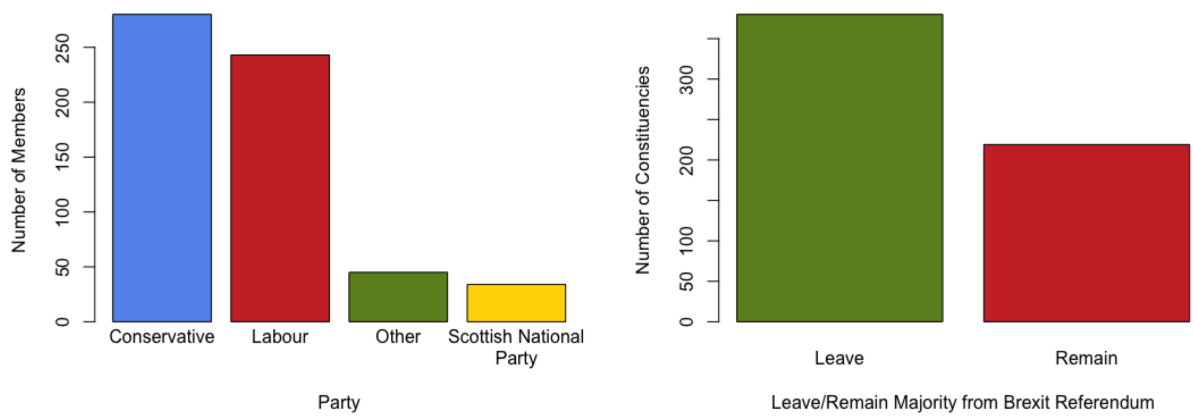

Figure 1: Caption