Political Position Estimation of Politicians using Social Media Communication

Karishma Rao, Master of Science in Computer Science
University of Dublin, Trinity College Dublin, 2022

Supervisor: Dr. Carl Vogel

Politicians make use of social media to communicate with the public and voice their opinions on different subject matters of public interest. Social media communication is an accessible and easy way for the politicians to express their stance and try and gain the public's favour. However, does this communication accurately represent the voting behaviour of the elected representatives? This project aims to use Twitter communications of the members of the Dáil Éireann (TDs or deputies), the lower house of the Oireachtas (the Irish Parliament) to try and predict their voting behaviour in the motions put forward in front of the Dáil Éireann. The information the general public is exposed to is limited to communications by the officials through different media. By trying to align the online text contributed by the elected officials with their behavior during legislative vote, it is possible to try to get a measure on whether there is delivery on promises made. As part of this study, the tweets of the members of the Dáil Éireann were extracted using Twitter API and the voting data of the deputies were extracted from the Oireachtas's official website. The two datasets were aligned and after text pre-processing and feature engineering techniques were employed, the data was evaluated, and predictions were made using three different classifiers. Another dataset was prepared that used only the voting behaviour of the TDs obtained from the Oireachtas's website and the models were evaluated against this dataset. The performance of twitter and voting data combined dataset models were compared against the models that used only the TDs voting data. This was done to analyse if there was any improvement in the predictions of the TD voting behaviour after adding features from the Twitter data. It was found that out of the three classifiers, the performance of the Decision Tree classifier significantly improved after aligning the twitter data with the individual deputy voting data. The Random Forest classifier performed the best in both cases. The results obtained in this study indicate that with a more finely tuned dataset, twitter data can be used to predict the voting behaviour of the TDs in the legislature.