The Library

Train on APC-2016-part1.txt has 8 chapters.

Test on APC-2016-part2.txt has 10 chapters.

Train on CPP-2016-part1.txt has 26 chapters.

Test on CPP-2016-part2.txt has 25 chapters.

Train on NDC-2016-part1.txt has 17 chapters.

Test on NDC-2016-part2.txt has 21 chapters.

Train on NPP-2016-part1.txt has 9 chapters.

Test on NPP-2016-part2.txt has 9 chapters.

Train on PPP-2016-part1.txt has 3 chapters.

Test on PPP-2016-part2.txt has 2 chapters.

**Confusion Matrix**

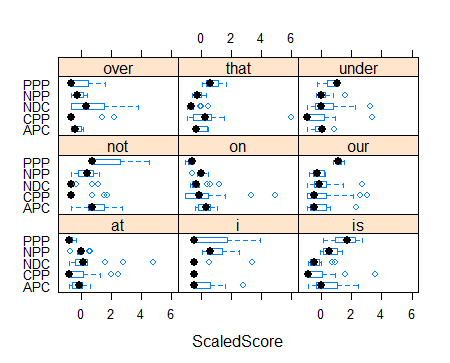
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrx  Actual/Predicted | APC | CPP | NDC | NPP | PPP | Error Rate |
| APC | 0 | 5 | 2 | 1 | 0 | 1.00 |
| CPP | 1 | 23 | 2 | 0 | 0 | 0.12 |
| NDC | 0 | 6 | 11 | 0 | 0 | 0.35 |
| NPP | 0 | 0 | 2 | 7 | 0 | 0.22 |
| PPP | 0 | 2 | 1 | 0 | 0 | 1.00 |

After training the randomForest model with the part1 (first parts) of the manifestos of all parties, it was noticed that 23 chapters out of 26 of CPP’s manifesto matched CPP’s writing style while 1 matched that of APC and 2 matched that of NDC. NPP also had 7 out of 9 chapters of its manifesto matching NPP’s writing style and the rest 2 matching NPP’s writing style. NDC had 11 out of 17 chapters of its manifesto matching its own writing style but the rest 6 chapters matches that of CPP. The model could not distinguish the writing styles of APC and PPP and this could be because the model had fewer chapters to train on. This shows that the randomForest could not clearly or fully distinguish the writing style of a specific party, although it did a good work at distinguishing the writing styles of CPP, NDC and NPP.

**The OOB Error Rate**

OOB estimate of error rate: 34.92%. This is quite high because it is near an average point and indicates that the model struggled to clearly distinguish between the writing styles of the various parties..

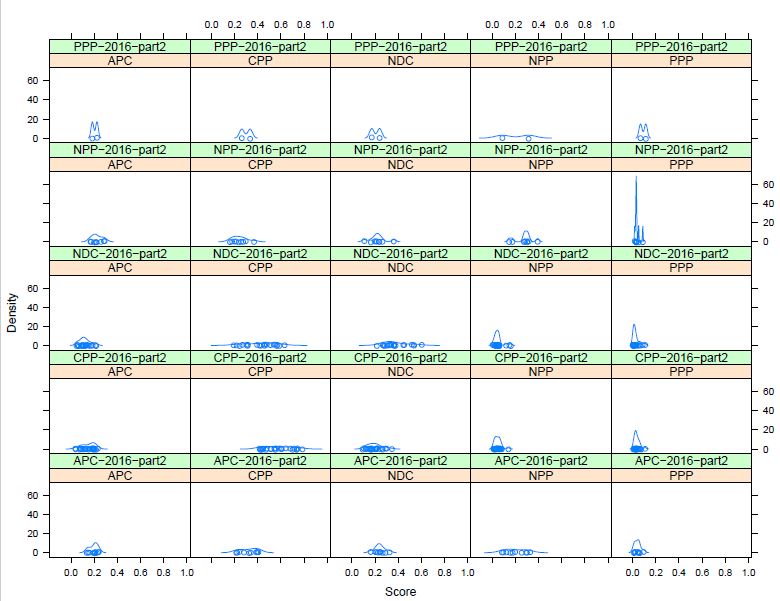
The graph on below shows those stop words which are most useful in distinguishing manifestos from each other. We see that the NDC Manifesto uses the words “over”, “at” and “under” more frequently than any other party. The words “not” and “I” are most frequently used by the PPP. Finally, the words “our” and “is” are most frequently used by the PPP.



**Guess**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | APC | CPP | NDC | NPP | PPP |
| APC | 0 | 7 | 0 | 3 | 0 |
| CPP | 0 | 25 | 0 | 0 | 0 |
| NDC | 0 | 15 | 6 | 0 | 0 |
| NPP | 1 | 2 | 1 | 5 | 0 |
| PPP | 0 | 1 | 0 | 1 | 0 |

The random forest could not predict any of the chapters for APC. For PPP, the random forest could not also predict any of the chapters. 15 chapters of NDC Manifesto matches CPP writing style and 2 chapters of NPP Manifesto matches CPP writing style. 3 chapters of APC Manifesto matches NPP writing styles. 1 chapter of PPP Manifesto matches CPP and NPP writing styles. The model could correctly predict CPP’s writing style since all 25 chapters of CPP’s manifesto matches that of CPP’s writing style. The model also did quite a good work in predicting NPP’s writing style, looking at the fact that 5 out of 9 chapters tested matches that of NPP’s writing style, while 1, 2 and 1 matched NDC, CPP and APC respectively. Hence, it can be concluded that the randomForest model was able to predict the writing style of just two parties (CPP and NPP) out of the manifestos of 5 parties.



**Interpretation**

For most of the manifestos analyzed by the program, the number of common words between them were less than twenty hence it is safe to assume that they were driving at different messages. However, for PPP-2016-part2 and APC, PPP-2016-part2 and PPP, NDC-2016-part2 and PPP, CPP-2016-part2 and PPP had a commonality of more than 20 words whiles NPP-2016-part2 and PPP manifestos had over 60 words in common. Thus, it can safely be assumed that the above stated manifestos were driving at similar interest and policies.