The Library

APC-2016-part1.txt has 8 chapters.

APC-2016-part2.txt has 10 chapters.

CPP-2016-part1.txt has 26 chapters.

CPP-2016-part2.txt has 25 chapters.

NDC-2016-part1.txt has 17 chapters.

NDC-2016-part2.txt has 21 chapters.

NPP-2016-part1.txt has 9 chapters.

NPP-2016-part2.txt has 9 chapters.

PPP-2016-part1.txt has 3 chapters.

PPP-2016-part2.txt has 2 chapters.

**Confusion Matrix**

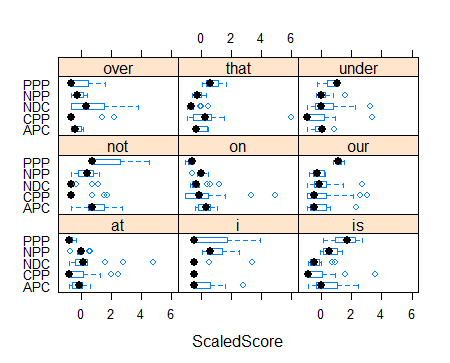
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Confusion Matrx  Actual/Predicted | APC | CPP | NDC | NPP | PPP | Error Rate |
| APC | 0 | 5 | 2 | 1 | 0 | 1.00 |
| CPP | 1 | 23 | 2 | 0 | 0 | 0.12 |
| NDC | 0 | 6 | 11 | 0 | 0 | 0.35 |
| NPP | 0 | 0 | 2 | 7 | 0 | 0.22 |
| PPP | 0 |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

After training and testing our random forest, it was unable to clearly distinguish each party’s manifesto. The random forest could not identify the first section (actual chapters) of the APC Manifesto against its second section (predicted chapters). The PPP Manifesto also suffers the same fate. No section of all the Parties’ Manifestos matches that of the CPP Manifesto.

**The OOB Error Rate**

OOB estimate of error rate: 34.92%. This is quite high because it is near an average point.

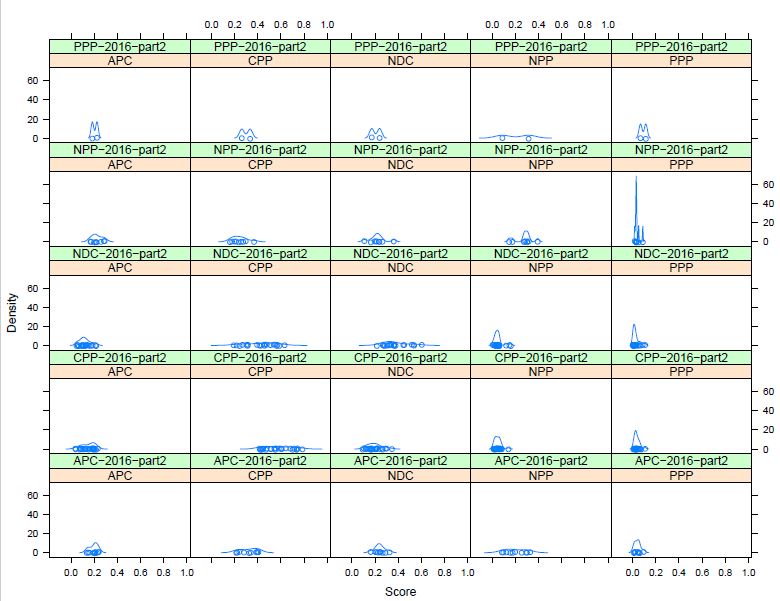
The graph on the below shows those stop words which are most useful in distinguishing manifestos from each other. We see that the NDC Manifesto uses the words “over”, “at” and “under” more frequently than any other party. The words “not” and “I” are most frequently used by the PPP. Finally, the words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



**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.